

# Moving target inference with hierarchical Bayesian models in synthetic aperture radar imagery

Gregory E. Newstadt, *Student Member, IEEE*, Edmund G. Zelnio, and Alfred O. Hero III, *Fellow, IEEE*

February 20, 2013

## Abstract

In synthetic aperture radar (SAR), images are formed by focusing the response of stationary objects to a single spatial location. On the other hand, moving targets cause phase errors in the standard formation of SAR images that cause displacement and defocusing effects. SAR imagery also contains significant sources of non-stationary spatially-varying noises, including antenna gain discrepancies, angular scintillation (glints) and complex speckle. In order to account for this intricate phenomenology, this work combines the knowledge of the physical, kinematic, and statistical properties of SAR imaging into a single unified Bayesian structure that simultaneously (a) estimates the nuisance parameters such as clutter distributions and antenna miscalibrations and (b) estimates the target signature required for detection/inference of the target state. Moreover, we provide a Monte Carlo estimate of the posterior distribution for the target state and nuisance parameters that infers the parameters of the model directly from the data, largely eliminating tuning of algorithm parameters. We demonstrate that our algorithm competes at least as well on a synthetic dataset as state-of-the-art algorithms for estimating sparse signals. Finally, performance analysis on a measured dataset demonstrates that the proposed algorithm is robust at detecting/estimating targets over a wide area and performs at least as well as popular algorithms for SAR moving target detection.

## I. INTRODUCTION

This work proposes algorithms for detecting and estimating targets in synthetic aperture radar (SAR) images. The image formation process for SAR images is more complicated than that of standard electro-optical images. Examples of these complexities include:

- SAR images have complex-valued rather than real-valued intensities, and the SAR phase information is of great importance for detection and estimation of target states. [1]–[3].
- SAR images are corrupted by spatiotemporally-varying antenna gain/phase patterns that often need to be estimated from homogeneous target-free data [4], [5].
- SAR images have spatially-varying clutter that can mask the target signature unless known a priori or properly estimated [6].
- SAR images contain motion-induced displacement and diffusion of the target response [1], [7].
- SAR images include multiple error sources due to radar collection and physical properties of the reflectors, such as angular scintillation (a.k.a. glints) [8] and speckle [9], [10].

Despite these complications, a great deal of structure exists in SAR images that can be leveraged to provide stronger SAR detection and tracking performance. This includes (a) using the coherence between multiple channels of an along-track radar in order to remove the stationary background (a.k.a, ‘clutter’), (b) assuming that pixels within the image can be described by one (or a mixture) of a small number of object classes (e.g., buildings, vegetation, etc.), and (c) considering kinematic models for the target motion such as Markov smoothness priors. From this structure in SAR imagery, one might consider models that

Gregory Newstadt and Alfred Hero are with the Dept. of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor. Edmund Zelnio is with the Air Force Research Laboratory, Wright Patterson Air Force Base, OH 45433, USA. E-mail: ({newstage},{hero})@umich.edu and edmund.zelnio@wpafb.af.mil).

The research in this paper was partially supported by Air Force Office of Scientific Research award FA9550-06-1-0324 and by Air Force Research Laboratory award FA8650-07-D-1221-TO1.

This document was cleared for public release under document number 88ABW-2013-0611.

assume that the clutter lies in a low-dimensional subspace that can be estimated directly from the data. Indeed, recent work Borcea et al. [11] has shown that SAR signals can be represented as a composition of a low-rank component containing the clutter, a sparse component containing the target signatures, and additive noise.

In general, SAR images are formed by focusing the response of stationary objects to a single spatial location. Moving targets, however, will cause phase errors in the standard formation of SAR images that cause displacement and defocusing effects. Most methods designed to detect the target depend on either (a) exploiting the phase errors induced by the SAR image formation process for a single phase center system or (b) canceling the clutter background using a multiple phase center system. In this chapter, we provide a rich model that can combine (and exploit) both sources of information in order to improve on both methodologies.

Fienup [1] provides an analysis of SAR phase errors induced by translational motions for single-look SAR imagery. He shows that the major concerns are (a) azimuth translation errors from range-velocities, (b) azimuth smearing errors due to accelerations in range, and (c) azimuth smearing due to velocities in azimuth. Fienup also provides an algorithm for detecting targets by their induced phase errors. The algorithm is based on estimating the moving target's phase error, applying a focusing filter, and evaluating the sharpness ratio as a detection statistic. Jao [7] shows that given both the radar trajectory and the target trajectory, it is possible to geometrically determine the location of the target signature in a reconstructed SAR image. Although the radar trajectory is usually known with some accuracy, the target trajectory is unknown. On the other hand, if the target is assumed to have no accelerations, Jao provides an efficient FFT-based method for refocusing a SAR image over a selection of range velocities. Khwaja and Ma [12] provide a algorithm to exploit the sparsity of moving targets within SAR imagery; they propose a basis that is constructed from trajectories formed from all possible combinations of a set of velocities and positions. To combat the computational complexity of searching through this dictionary, the authors use compressed sensing techniques. Instead of searching over a dictionary of velocities, our work proposes to use a prior distribution on the target trajectory that can be provided a priori through road and traffic models or adaptively through observations of the scene over time.

The process of removing the stationary background in order to detect moving targets is also known in the literature as 'change detection' or 'clutter suppression'. Gierull [13] provides a statistical analysis of the phase and magnitude of complex SAR images for two channels. He shows that SAR images cannot be modeled as spatially-invariant Gaussian in many cases of interest, such as in urban environments, where the statistics vary spatially and may be modulated by random variations. In our work, we model the distributions of the clutter as spatially varying and model the random modulations directly.

Ender [6] applies space-time adaptive processing (STAP) to multiple-channel SAR imagery. Similar to standard change detection algorithms such as displaced phase center array (DPCA) and along-track interferometry (ATI), STAP models the clutter as being embedded in a one-dimensional subspace. However, STAP extends those algorithms to using  $N > 2$  channels, where a single channel is used to estimate the stationary background and the remaining  $(N - 1)$  channels are used to estimate the moving component. However, STAP relies on estimating the complex-valued covariance matrix of the  $N$ -channel system, which in turn depends on the availability of homogeneous target-free secondary data.

There are a multitude of algorithms for change detection that are based on multi-temporal SAR images rather than multi-channel data. Bazi and Bruzzone [14] develop methods for multi-temporal change detection that use adaptive thresholds for declaring changes based on a theoretical analysis of a generalized Gaussian model. Bovolo and Bruzzone [15] provide another algorithm for change detection that employs a wavelet-based multiple scale decomposition of multitemporal SAR images, with an adaptive scale driven fusion algorithm.

Ranney and Soumekh [4], [5] develop methods for change detection from SAR images collected at two distinct times that are robust to errors in the SAR imaging process. They address error sources including inaccurate position information, varying antenna gains, and autofocus errors. They propose that the stationary components of multi-temporal SAR images can be related by a spatially-varying 2-

dimensional filter. To make the change detection algorithm numerically practical, the authors propose that this filter can be well-approximated by a spatially invariant response within small subregions about any pixel in the image. This thesis adopts this model for the case where there are no registration errors. Under a Gaussian assumption for the measurement errors, it can be shown that the maximum likelihood estimate for the filter coefficients can be computed easily through simple least squares.

Ground Moving Target Indication (GMTI) methods involve the processing of SAR imagery to detect and estimate moving targets. Often clutter cancellation and change detection play a preprocessing role in these algorithms [16]–[19]. This chapter aims to combine properties of many of these algorithms into a unifying framework that simultaneously estimates the target signature and the nuisance parameters, such as clutter distributions and antenna calibrations.

It should be noted that many of the previously discussed algorithms work well in certain situations, but do not provide estimates of their uncertainty that may be necessary for adaptive sensing, sensor management, or sensor fusion. This chapter aims to bridge this gap by providing a Bayesian formulation that provides uncertainty distributions for the presence of the moving targets and their positions. Under this Bayesian formulation, we can generate the posterior distribution of the target state(s) given the observations (i.e., the SAR images).

Recently, there has been great interest by Wright et al. [20], Lin et al. [21], Candes et al. [22] and Ding et al. [23] in the so-called robust principal component analysis (RPCA) problem that decomposes high-dimensional signals as

$$\mathbf{I} = \mathbf{L} + \mathbf{S} + \mathbf{E}, \quad (1)$$

where  $\mathbf{I} \in \mathbb{R}^{N \times M}$  is an observed high dimensional signal,  $\mathbf{L} \in \mathbb{R}^{N \times M}$  is a low-rank matrix with rank  $r \ll NM$ ,  $\mathbf{S} \in \mathbb{R}^{N \times M}$  is a sparse component, and  $\mathbf{E} \in \mathbb{R}^{N \times M}$  is dense low-amplitude noise. In [20]–[22], inference in this model is done by optimizing a cost function of the form

$$\arg \min_{\mathbf{L}, \mathbf{S}} \|\mathbf{L}\|_* + \gamma \|\mathbf{S}\|_1 + (2\mu)^{-1} \|\mathbf{I} - \mathbf{L} - \mathbf{S}\|_F \quad (2)$$

where the last term is sometimes replaced by the constraint  $\mathbf{I} = \mathbf{L} + \mathbf{S}$ . One major drawback of these methods involves finding the algorithm parameters (e.g., tolerance levels or choices of  $\gamma, \mu$ ), which may depend on the given signal. Moreover, it has been demonstrated that the performance of these algorithms can depend strongly on these parameters.

Bayesian methods by Ding et al. [23] have been proposed that simultaneously learn the noise statistics and infer the low-rank and sparse components. Moreover, they show that their method can be generalized to richer models, e.g. Markov dependencies on the target locations. Additionally, these Bayesian inferences provide a characterization of the uncertainty of the outputs through a Markov Chain Monte Carlo (MCMC) estimate of the posterior distribution.

The work by Ding et al. [23] is based on a general Bayesian framework [24] by Tipping for obtaining *sparse* solutions to regression and classification problems. Tipping’s framework uses simple distributions (e.g., those belonging to the exponential class) that can be described by few parameters, known as hyperparameters. Moreover, Tipping considers a *hierarchy* where the hyperparameters themselves are assumed to have a known ‘hyperprior’ distribution. Often the prior and hyperprior distributions are chosen to be conjugate, so that inference is simple. Tipping provides insight into choosing the hyperparameter distributions so as to be non-informative with respect to the prior. This latter property is important in making it possible to implement inference algorithms with few tuning parameters. Finally, Tipping provides a specialization to the ‘relevance vector machine’ (RVM), which can be thought of as a Bayesian version of the support vector machine. Wipf et al. [25] provides an interpretation of the RVM as the application of a variational approximation to estimating the true posterior distribution. Wipf et al. explains the sparsity properties of the sparse Bayesian learning algorithms in a rigorous manner. Additionally, it also provides connections with other popular work in sparse problems, such as the FOCUSS and basis pursuit algorithms.

We adopt this hierarchical Bayesian model to SAR images. This requires the following non-trivial extensions: (a) we consider complex-valued data rather than real-valued intensity images; (b) we model

correlated noise sources based on physical knowledge of SAR phase history collection and image formation; (c) we relax the assumption of a low-rank background component by assuming that the background component lies in a low-dimensional subspace; and (d), we directly model SAR phenomenology by including terms for glints, speckle contributions, antenna gain patterns, and target kinematics. Moreover, we demonstrate the performance of the proposed algorithm on both simulated and measured datasets, showing competitive or better performance in a variety of situations.

The rest of the paper is organized as follows: Notation is given in Section II and the image model is provided in III. Markov, spatial, and/or target kinematic extensions are discussed in Section IV. The inference algorithm is given in Section V. Performance is analyzed over both simulated and measured datasets in Section VI. We conclude and point to future work in Section VII.

## II. NOTATION

Available is a set of SAR images of a region formed from multiple passes of an along-track radar platform with multiple antennas (i.e., phase centers.) Moreover, images are formed over distinct azimuth angle ranges that can be indexed by the frame number,  $f$ . Table I provides the indexing scheme used throughout this chapter in order to distinguish between images from various antennas, frames, and/or passes. Table II provides a list of indexing conventions used to denote collections of variables.

TABLE I  
INDEX VARIABLE NAMES USED IN PAPER

Index Description	Index Variable	Range
Antenna (channel)	$k$	$1, 2, \dots, K$
Frame (azimuth range)	$f$	$1, 2, \dots, F$
Pass	$i$	$1, 2, \dots, N$
Pixel	$p$	$1, 2, \dots, P$

TABLE II  
OUR DATA INDEXING CONVENTIONS

Variable	Convention	Description
$i_{k,f,i}^{(p)}$	Standard	Value at pixel $p$ , antenna $k$ , and frame $f$ , pass $i$
$\underline{i}_{f,i}^{(p)}$	Underline	Values at pixel $p$ , frame $f$ , and pass $i$ over all antennas
$\boldsymbol{i}_{f,1:N}^{(p)}$	Lower-case, Boldface	Values at pixel $p$ and frame $f$ over all antennas and passes
$\boldsymbol{I}_{f,i}$	Upper-case Boldface	Values over all pixels and antennas at frame $f$ and pass $i$
$\boldsymbol{I}$	Upper-case, Boldface, No Indices	Values over all pixels, antennas, frames, and passes

We model the quadrature components of the SAR images with the complex-normal distribution, where we use the notation

$$\underline{w} \sim \mathcal{CN}(0, \Gamma) \quad (3)$$

where  $\mathcal{CN}(\mu, \Gamma)$  represents the complex-Normal distribution with mean  $\mu$  and complex covariance matrix  $\Gamma$ , and  $\vec{w}$  is random vector of  $K$  complex-values (from each of  $K$  antennas.)

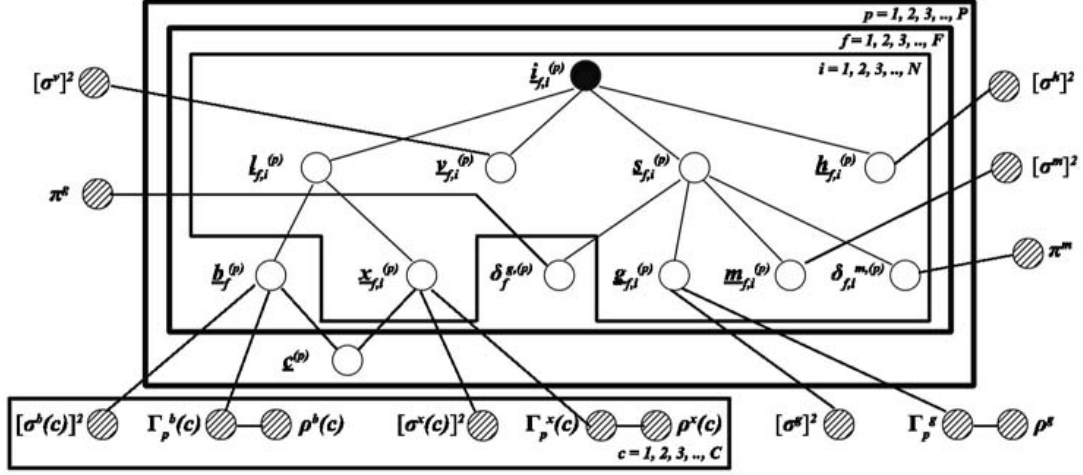


Fig. 1. This figure provides a graphical representation of the proposed SAR image model. The dark circle represents the observed random variable. The unshaded circles represent the basic parameters of the model, while the dashed circles represent hyperparameters that are also modeled as random variables.

### III. SAR IMAGE MODEL

We propose a decomposition of SAR images at each frame  $f$  and pass  $i$  as follows

$$I_{f,i} = H_{f,i} \circ (L_{f,i} + S_{f,i} + V_{f,i}), \quad (4)$$

where  $H_{f,i}$  is a spatiotemporally-varying filter that accounts for antenna calibration errors,  $L_{f,i}$  is a low-dimensional representation of the background clutter,  $S_{f,i}$  is a sparse component that contains the targets of interest,  $V_{f,i}$  is zero-mean additive noise, and  $\circ$  denotes the Hadamard (element-wise) product. Each of these components belongs to the space  $\mathbb{C}^{P \times K}$ . The remainder of this section discusses the model in detail. Figure III shows a graphical representation of the model.

#### A. Low-dimensional component, $L_{f,i}$

We propose a decomposition of the low-rank component as

$$L_{f,i} = B_f + X_{f,i}, \quad (5)$$

where  $B_f$  is the inherent background that is identical over all passes,  $X_{f,i}$  is the speckle noise component that arises from coherent imaging in SAR. Posner [9] and Raney [10] describe speckle noise, which tends to be spatially correlated depending on the texture of the surrounding pixels.

The quadrature components of radar channels are often modeled as zero-mean Gaussian processes, though Gierull [13] demonstrates that for heterogeneous clutter (such as in urban scenes), one must consider spatially varying models. To account for this spatial variation, this model assumes that each background pixel can be defined by one of  $J$  classes that may be representative of roads, vegetation, or buildings within the scene. Our model is low-dimensional since  $J \ll P$ , where  $P$  is the number of pixels in the measured images. We put a multinomial model on each object class

$$\underline{c}^{(p)} = \left\{ c_j^{(p)} \right\}_{j=1}^J \sim \text{Multinomial}(1; q_1, q_2, \dots, q_J) \quad (6)$$

where  $q_j$  is the prior probability of the  $j$ -th object class. Then the class assignment  $C^{(p)}$  is the single location in  $\underline{c}$  with value equal to one. We use a hidden Markov model dependency that reflects that neighboring pixels are likely to have the same class. The class  $C^{(p)}$  defines the distribution of the pixel

$p$ , where we specifically model the background and speckle components respectively as complex-normal distributed:

$$\underline{b}_f^{(p)} \sim \mathcal{CN}(0, \mathbf{\Gamma}_B^{C(p)}), \quad \underline{x}_{f,i}^{(p)} \sim \mathcal{CN}(0, \mathbf{\Gamma}_X^{C(p)}) \quad (7)$$

Note that the class type specifies the distribution of the pixels and each vector of  $K$  values (e.g. background  $\underline{b}_f^{(p)}$  or speckle  $\underline{x}_{f,i}^{(p)}$ ) is drawn independently from that distribution.

### B. Sparse component, $\mathbf{S}_{f,i}$

The sparse component is modeled as

$$\mathbf{S}_{f,i} = (\mathbf{\Delta}_f^G \otimes \mathbf{1}_K^T) \circ \mathbf{G}_{f,i} + (\mathbf{\Delta}_{f,i}^M \otimes \mathbf{1}_K^T) \circ \mathbf{M}_{f,i}, \quad (8)$$

where  $\mathbf{G}_{f,i} \in \mathbb{C}^{P \times K}$  is the specular noise (glints) component with associated indicator variables  $\mathbf{\Delta}_f^G \in \{0, 1\}^P$ ,  $\mathbf{M}_{f,i} \in \mathbb{C}^{P \times K}$  is the (moving) target component with associated indicator variables  $\mathbf{\Delta}_{f,i}^M \in \{0, 1\}^P$ ,  $\mathbf{1}_K$  is the all ones vector of size  $K \times 1$ , and  $\otimes$  is the Kronecker product. Note that this shared sparsity model assumes that the glint/target components are present in one antenna if and only if they are present in the other antennas. Moreover, glints are known to have a large angular dependence, in the sense that the intensity of the glint dominates in only a few azimuth angles but is present from pass to pass as described by Borden [8]. Thus, the indicators for glints do not depend on the pass index  $i$ . Once again, we assume that the glints and target components are zero-mean complex-normal distributed with covariances  $\mathbf{\Gamma}_G$  and  $\mathbf{\Gamma}_M$ , respectively.

The indicator variable  $\delta^{z,(p)}$  at pixel  $p$  where  $z$  is representative of either  $g$  or  $m$  is modeled as

$$\delta^{z,(p)} \sim \text{Bernoulli}(\pi^{z,(p)}), \quad (9)$$

$$\pi^{z,(p)} \sim \text{Beta}(a_\pi, b_\pi) \quad (10)$$

A sparseness prior is obtained by setting  $a_\pi/[a_\pi + b_\pi] \ll 1$ . Alternatively, we can introduce additional structure in our model by letting  $a_\pi$  and  $b_\pi$  depend on previous frames (temporally) and/or neighboring pixels (spatially). This is particularly useful for detecting multi-pixel targets that move smoothly through a scene. Section IV discusses this modification in greater detail.

### C. Distribution of quadrature components

Many SAR detection algorithms rely on the ability to separate the target from the background clutter by assuming that the clutter lies in a low-dimensional subspace of the data. Consider a random vector of complex variables  $\underline{w} \sim \mathcal{CN}(0, \mathbf{\Gamma})$  where  $w$  is representative of  $b$ ,  $x$ ,  $g$  or  $m$ . Under the assumptions that (a) the quadrature components of each antenna are zero-mean normal with variance  $\sigma^2$  and (b) the correlation among components  $w_m$  and  $w_n$  is given by  $\rho e^{-j\phi_{mn}}$ , then  $\mathbf{\Gamma}$  can be shown to have the form

$$\mathbf{\Gamma} = \sigma^2 \begin{bmatrix} 1 & \rho e^{-j\phi_{12}} & \dots & \rho e^{-j\phi_{1K}} \\ \rho e^{j\phi_{12}} & 1 & \dots & \rho e^{-j\phi_{2K}} \\ \vdots & \vdots & \ddots & \vdots \\ \rho e^{j\phi_{1K}} & \rho e^{j\phi_{2K}} & \dots & 1 \end{bmatrix}, \quad (11)$$

where  $\sigma^2$  is the channel variance,  $\rho$  is the coherence between antennas, and  $\{\phi_{nm}\}_{n,m}$  are the interferometric phase differences between the antennas<sup>1</sup>. In an idealized model with a single point target, the interferometric phases  $\phi_{mn}$  can be shown to be proportional to the target radial velocity. In images containing only stationary targets (i.e., the background components), the covariance matrix has a simpler form:

$$\mathbf{\Gamma}_{\text{background}} = \sigma^2 ((1 - \rho) \mathbf{I}_{K \times K} + \rho \mathbf{1}_K \mathbf{1}_K^T) \quad (12)$$

<sup>1</sup>A more general model could account for different channel variance and coherence values, but since we use the calibration constants  $\mathbf{H}_{f,i}$  to equalize the channels, the effect was seen to be relatively insignificant.

where  $\mathbf{I}_{K \times K}$  is the  $K \times K$  identity matrix and  $\mathbf{1}_K$  is the all-ones vector of length  $K$ .

It should be noted that the covariance matrix in equation (11) is related directly to some common methods for change detection in SAR imagery. In particular, consider the two antenna case ( $K = 2$ ). Along-track interferometry (ATI) thresholds the phase  $\phi_{12}$  in order to detect moving targets which have non-zero phases. Moreover, one can easily show that the eigendecomposition of  $\Gamma$  leads to eigenvalues  $\lambda$  and eigenvectors  $\nu$ :

$$\lambda(\Gamma) = \{2\sigma^2(1 + \rho), 2\sigma^2(1 - \rho)\} \quad (13)$$

$$\nu(\Gamma) = \left\{ \begin{bmatrix} 1 \\ e^{-j\phi_{12}} \end{bmatrix}, \begin{bmatrix} 1 \\ -e^{-j\phi_{12}} \end{bmatrix} \right\}. \quad (14)$$

Displaced phase center array (DPCA) processing thresholds the difference between the two channels. Indeed, for small phases, the second eigenvector of  $\Gamma$  reduces to  $[1; -1]^T$ . Thus DPCA can be interpreted as a projection onto the eigenvector of  $\Gamma$ . Deming [2] shows that ATI performs well when canceling bright clutter (i.e., high  $\sigma^2$  and  $\rho \approx 1$ ), while DPCA performs well for canceling dim clutter (i.e., small  $\sigma^2$  and  $\rho \approx 0$ .) In our work, we combine the discriminating power of both DPCA and ATI by modeling the covariance matrices directly. Ender [6] provides space-time adaptive processing (STAP), where optimal detection schemes for moving targets are based on the estimation of  $\Gamma$ . However, the performance of STAP depends on the availability of target-free homogeneously distributed measurements in order to estimate  $\Gamma$  effectively. In this chapter, we simultaneously estimate the covariance matrices as well as the target contributions. Thus, we demonstrate the capability to detect targets even in the presence of heterogeneous measurements.

In this thesis, the covariance matrix  $\Gamma$  is modeled as a random variable using a modified version of the Multivariate-Normal-Inverse-Wishart conjugate distributions. In particular, we let

$$\underline{w} \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \Gamma_\rho) \quad (15)$$

$$\Gamma_\rho \sim \text{InvWishart}(a_\Gamma((1 - \rho)\mathbf{I}_{K \times K} + \rho\mathbf{1}_K\mathbf{1}_K^T), \nu_\Gamma) \quad (16)$$

$$\sigma^2 \sim \text{InvGamma}(a_\sigma, b_\sigma) \quad (17)$$

$$\rho \sim \text{Beta}(a_\rho, b_\rho) \quad (18)$$

where  $a_\sigma = b_\sigma = 10^{-6}$  as suggested by Tipping [24] to promote non-informative priors,  $(a_\rho, b_\rho)$  are chosen so that  $\rho \approx 1$  to ensure a high coherence among the background components,  $\nu_\Gamma$  is a parameter that controls how strongly to weight the prior covariance matrix, and  $a_\Gamma$  is chosen so that  $E[\Gamma_\rho] = (1 - \rho)\mathbf{I}_{K \times K} + \rho\mathbf{1}_K\mathbf{1}_K^T$ . In this thesis, we choose  $\nu_\Gamma$  to be large to reflect our belief that  $\sigma^2 \Gamma_\rho$  should be close to equation (11). Note that this model separates the learning of the channel variance  $\sigma^2$ , which we have no a priori knowledge about, from the learning of the correlation structure  $\Gamma_\rho$ .

#### D. Calibration filter, $\mathbf{H}_{f,i}$

The calibration constants are assumed to be constant within small spatial regions  $p \in Z_g$ , though they may vary as a function of antenna, frame, or pass. In particular, we let

$$h_{k,f,i}^{(p)} = z_{k,f,i}(g), \forall p \in Z_g, \quad (19)$$

$$z_{k,f,i}(g) \sim \mathcal{CN}(1, (\sigma^H)^2) \quad (20)$$

where we note that if  $(\sigma^H)^2$  is large, then maximum likelihood inference in this case yields the least-squares solution.

### E. Summary of SAR Image Model

Tables III provides a summary of the distributions for the proposed decomposition of SAR images. The table also provides a characterization of spatial (across pixels) and temporal (across frames and passes) dependencies. For example, background and speckle components have distributions characterized by their class  $j$ . Thus, all pixels with class  $j$  belong to a subset  $Q_j \subset \{1, 2, \dots, P\}$ . In contrast, the distribution of moving targets is assumed to be identical across all pixels, yet the distribution of their indicators varies for each pixel, frame, and pass.

Tables IV and V provide a summary of the parameters of the distributions in Table III. We provide the simple model for target and glint indicator probabilities that just assumes that they are sparse in the image. We can introduce additional richness in the model by allowing the parameters  $a_\pi$  and  $b_\pi$  to vary over pixels, frames, and passes as described in Section IV-A.

## IV. MARKOV/SPATIAL/KINEMATIC MODELS FOR THE SPARSE COMPONENT

### A. Indicator probability models

This model contains multiple indicator variables with prior probabilities distributed as  $\text{Beta}(a_\pi, b_\pi)$ . Moreover, sparsity is obtained when  $a_\pi/[a_\pi + b_\pi] \ll 1$ . Alternatively, we can introduce additional structure in our model by letting  $a_\pi$  and  $b_\pi$  depend on previous frames (temporally) and/or neighboring pixels (spatially). This is especially useful for detecting multi-pixel targets that move smoothly through a scene.

Define  $W^M(p, \Delta_{f,i}^M)$  to be a function that maps the indicator variables  $\Delta_{f,i}^M$  to a real number. For example, this may be the average number of non-zero indicators in the neighborhood of pixel  $p$ , or a weighted version that puts higher value on neighboring pixels. For  $f = 1$ , we let

$$\begin{bmatrix} a_{1,i}^M(p) \\ b_{1,i}^M(p) \end{bmatrix} = \begin{cases} [a_H & b_H]^T, & W^M(p, \Delta_{1,i}^M) > \varepsilon_{\text{spatial}}^M, \\ [a_L & b_L]^T, & \text{else,} \end{cases} \quad (21)$$

and for  $f > 1$

$$\begin{bmatrix} a_{f,i}^M(p) \\ b_{f,i}^M(p) \end{bmatrix} = \begin{cases} [a_H & b_H]^T, & W^M(p, \Delta_{f,i}^M) > \varepsilon_{\text{spatial}}^M \text{ and} \\ & W^M(p, \Delta_{f-1,i}^M) > \varepsilon_{\text{temporal}}^M, \\ [a_L & b_L]^T, & \text{else.} \end{cases} \quad (22)$$

In this chapter, we choose  $(a_L, b_L, a_H, b_H)$  so that  $a_L/(a_L + b + L) \ll 1$  and  $a_H/(a_H + b + H) \gg 0$ . A similar model can be introduced for the probabilities of the glints.

### B. Target kinematic model

In some applications, such as target tracking or sequential detection, we may have access to an estimate of the kinematic state of the target(s) of interest, such as position, velocity and acceleration. This may be useful for predicting the location of the target at sequential frames. For simplicity, consider a single target at time  $\tau$  whose state  $\xi(\tau) = (\mathbf{r}(\tau), \dot{\mathbf{r}}(\tau))$  is known with standard errors  $\Sigma_\xi(\tau)$ . Note that the uncertainty model for  $(\mathbf{r}, \dot{\mathbf{r}})$  may be (a) known a prior from road maps or traffic behavior patterns, or (b) learned adaptively using some signal processing algorithm such as the Kalman or particle filters.

In standard SAR image formation, moving targets tend to appear displaced and defocused as described by Fienup [1] and Jao [7]. Moreover, Jao showed that given the radar trajectory  $(\mathbf{q}, \dot{\mathbf{q}})$  and the target trajectory  $(\mathbf{r}, \dot{\mathbf{r}})$ , one can predict the location of the target signature within the image  $\mathbf{p}$  by solving a system of equations that equate Doppler shifts and ranges, respectively, at each pulse:

$$\frac{d}{d\tau} [\|\mathbf{p} - \mathbf{q}(\tau)\|_2 - \|\mathbf{r}(\tau) - \mathbf{q}(\tau)\|_2]_{\mathbf{p}=\mathbf{p}^*} = 0 \quad (23)$$

$$\|\mathbf{p}^* - \mathbf{q}(\tau)\|_2 = \|\mathbf{r}(\tau) - \mathbf{q}(\tau)\|_2, \quad (24)$$



TABLE III  
DISTRIBUTIONAL MODELS FOR EACH COMPONENT IN EQUATIONS (4), (5), AND (8)

Component	Variable	Distribution	Parameters	Spatial	Temporal
Background	$\underline{b}_f^{(p)}$	$\mathcal{CN}(0, \Gamma^B(j))$	$\Gamma^B(j) = [\sigma^B(j)]^2 \mathbf{\Gamma}_\rho^B(j)$	$p \in Q_j$	All $f$ , All $i$
Speckle	$\underline{x}_{f,i}^{(p)}$	$\mathcal{CN}(0, \Gamma^X(j))$	$\Gamma^X(j) = [\sigma^X(j)]^2 \mathbf{\Gamma}_\rho^X(j)$	$p \in Q_j$	All $f$ , All $i$
Glints	$\underline{g}_{f,i}^{(p)}$	$\mathcal{CN}(0, \Gamma^G)$	$\Gamma^G = [\sigma^G]^2 \mathbf{\Gamma}_\rho^G$	All $p$	All $f$ , All $i$
Moving targets	$\underline{m}_{f,i}^{(p)}$	$\mathcal{CN}(0, \Gamma^M)$	$\Gamma^M = [\sigma^M]^2 \mathbf{I}_{K \times K}$	All $p$	All $f$ , All $i$
Additive noise	$V_{k,f,i}(p)$	$\mathcal{CN}(0, \Gamma^V)$	$\Gamma^V = [\sigma^V]^2 \mathbf{I}_{K \times K}$	All $p$	All $f$ , All $i$
Glint indicator	$\delta_f^{G,(p)}$	Bernoulli $\left(\pi_f^{G,(p)}\right)$	$\pi_f^{G,(p)}$	Each $p$	Each $f$ , All $i$
Target indicator	$\delta_{f,i}^{M,(p)}$	Bernoulli $\left(\pi_{f,i}^{M,(p)}\right)$	$\pi_{f,i}^{M,(p)}$	Each $p$	Each $f$ , Each $i$
Class assignment	$\underline{c}(p)$	Multinomial $(1; \underline{q})$	$\underline{q}$	Each $p$	All $f$ , All $i$
Calibration coefficients	$H_{k,f,i}(p) = z_{k,f,i}(g)$	$\mathcal{CN}(1, [\sigma^H]^2)$	$[\sigma^H]^2$	$p \in Z_g$	Each $f$ , Each $i$

TABLE IV  
DISTRIBUTIONAL MODELS FOR COVARIANCE PARAMETERS OF DISTRIBUTIONS IN TABLE III

Component	Variable	Distribution	Parameters	Suggested Value	Region
Background covariance	$\mathbf{\Gamma}^B(j) = [\sigma^B(j)]^2 \mathbf{\Gamma}_\rho^B(j)$	Inv-Gamma	$a_\sigma, b_\sigma$	$10^{-6}, 10^{-6}$	Each $Q_j$
Variance	$[\sigma^B(j)]^2$	Inv-Wishart	$a_\Gamma, \nu_\Gamma$	See note <sup>b</sup> , $O(P)$	
Correlation matrix	$\mathbf{\Gamma}_\rho^B(j)$	Beta	$a_\rho, b_\rho$	$a_\rho / (a_\rho + b_\rho) \approx 1$	
Coherence	$\rho^B(j)$				
Speckle covariance	$\mathbf{\Gamma}^X(j) = [\sigma^X(j)]^2 \mathbf{\Gamma}_\rho^X(j)$	Inv-Gamma	$a_\sigma, b_\sigma$	$10^{-6}, 10^{-6}$	Each $Q_j$
Variance	$[\sigma^X(j)]^2$	Inv-Wishart	$a_\Gamma, \nu_\Gamma$	See note <sup>b</sup> , $O(P)$	
Correlation matrix	$\mathbf{\Gamma}_\rho^X(j)$	Beta	$a_\rho, b_\rho$	$a_\rho / (a_\rho + b_\rho) \approx 1$	
Coherence	$\rho^X(j)$				
Glint covariance	$\mathbf{\Gamma}^G = [\sigma^G]^2 \mathbf{\Gamma}_\rho^G$	Inv-Gamma	$a_\sigma, b_\sigma$	$10^{-6}, 10^{-6}$	All $p$
Variance	$[\sigma^G]^2$	Inv-Wishart	$a_\Gamma, \nu_\Gamma$	See note <sup>b</sup> , $O(P)$	
Correlation matrix	$\mathbf{\Gamma}_\rho^G$	Beta	$a_\rho, b_\rho$	$a_\rho / (a_\rho + b_\rho) \approx 1$	
Coherence	$\rho^G$				
Target covariance	$\mathbf{\Gamma}^M = [\sigma^M]^2 \mathbf{I}_{K \times K}$	Inv-Gamma	$a_\sigma, b_\sigma$	$10^{-6}, 10^{-6}$	All $p$
Variance	$[\sigma^M]^2$				
Additive noise covariance	$\mathbf{\Gamma}^V = [\sigma^V]^2 \mathbf{I}_{K \times K}$	Inv-Gamma	$a_\sigma, b_\sigma$	$10^{-6}, 10^{-6}$	All $p$
Variance	$[\sigma^V]^2$				

TABLE V  
DISTRIBUTIONAL MODELS FOR OTHER PARAMETERS OF DISTRIBUTIONS IN TABLE III

Component	Variable	Distribution	Parameters	Suggested Value	Region
Variance	$[\sigma^V]^2$	Inv-Gamma	$a_\sigma, b_\sigma$	$10^{-6}, 10^{-6}$	
Target indicator probability	$\pi_{f,i}^M(p)$	Beta	$a_\pi, b_\pi$	$a_\pi / (a_\pi + b_\pi) \ll 1$	Each $p, f, i$
Glint indicator probability	$\pi_f^Y(p)$	Beta	$a_\pi, b_\pi$	$a_\pi / (a_\pi + b_\pi) \ll 1$	Each $p, f$
Region type probabilities	$\underline{q} = \{q_1, \dots, q_J\}$	Dirichlet	$\{e_j\}_{j=1}^J$	$e_j = 1/J$	All $p$
Calibration coefficient variance	$(\sigma^H)^2$	Inv-Gamma	$a_\sigma, b_\sigma$	$10^{-6}, 10^{-6}$	All $p$

```

procedure  $\{\Theta\}_{i=1:N_{samples}} = \text{SARGibbs}(\Theta_0, \mathbf{I})$ 
   $\Theta \leftarrow \Theta_0$ 
  for  $iteration = 1$  to  $N_{burnin} + N_{samples}$  do
    Sample  $\sim f(\mathbf{B}, \mathbf{X}, \mathbf{G}, \mathbf{M}, \Delta^G, \Delta^M | \mathbf{I}, -)$  //Base
    Sample  $\sim f(\mathbf{H} | \mathbf{I}, -)$  //Calibration filter
    Sample  $\sim f(\mathbf{C} | \mathbf{I}, -)$  //Class assignment
    Sample  $\sim f(\boldsymbol{\eta} | \mathbf{I}, -)$  //Hyper-parameters
     $\Theta_{iteration-N_{burnin}} \leftarrow \Theta$  if  $iteration > N_{burnin}$ 
  end for
end procedure

```

Fig. 2. Gibbs Sampling Pseudocode

which can be reduced to the simpler system of equations:

$$\dot{\mathbf{q}}(\tau) \cdot [\mathbf{p}^* - \mathbf{q}(\tau)] = [\dot{\mathbf{r}}(\tau) - \dot{\mathbf{q}}(\tau)] \cdot [\mathbf{r}(\tau) - \mathbf{q}(\tau)] \quad (25)$$

$$\|\mathbf{p}^* - \mathbf{q}(\tau)\|_2 = \|\mathbf{r}(\tau) - \mathbf{q}(\tau)\|_2 \quad (26)$$

The probable locations of the target can be predicted by one of several methods, including:

- Monte Carlo estimation of the target posterior density.
- Gaussian approximation using linearization or the unscented transformation to approximate the posterior density
- Analytical approximation.

Given an estimate of the posterior density, we can modify the function  $W^M$  described in the previous section to include dependence on this kinematic information. Details of the posterior density estimation are provided in Appendix A.

## V. INFERENCE

In the proposed hierarchical model, the distribution of hyper-parameters at the base layer are generally chosen to be conjugate to the distributions at the next layer. This allows for efficient approximation methods for the posterior distribution in the sense that we can sample exactly from these distributions. In particular, we use a Markov Chain Monte Carlo (MCMC) algorithm in the form of a Gibbs sampler to iteratively estimate the full joint posterior. In MCMC, this distribution is approximated by drawing samples iteratively from the conditional distribution of each (random) model variable given the most recent estimate of the rest of the variables (which we denote by  $-$ ). Let  $\Theta = \{\mathbf{B}, \mathbf{X}, \mathbf{G}, \mathbf{M}, \Delta^G, \Delta^M, \mathbf{H}, \mathbf{C}, \boldsymbol{\eta}\}$  represent a current estimate of all of the model variables where  $\boldsymbol{\eta}$  represents the set of all hyper-parameters. Given measurements  $\mathbf{I}$ , the inference algorithm is given in Figure 2. Note that MCMC algorithms require a burn-in period after the Markov chain has become stable, where the duration of burn-in period depends on the problem. After this point, we collect  $N_{samples}$  samples that represent the full joint distribution. However, we point out a couple of important features here. First, the sampling of the base model can be rewritten as

$$\begin{aligned}
 & f(\mathbf{B}, \mathbf{X}, \mathbf{G}, \mathbf{M}, \Delta^G, \Delta^M | \mathbf{I}, -) \\
 &= \prod_{p,f} f(\underline{b}_f^{(p)}, \mathbf{x}_{f,1:N}^{(p)}, \mathbf{g}_{f,1:N}^{(p)}, \mathbf{m}_{f,1:N}^{(p)}, \delta_f^{G,(p)}, \delta_{f,1:N}^{M,(p)} | \mathbf{I}, -)
 \end{aligned} \quad (27)$$

The conditional independence among pixels and frames given the nuisance parameters allows us to easily parallelize the sampling procedure over the largest dimensions of the state. Moreover, we can extend the

parallelization to sampling independently over passes by separating the sampling of equation (27) into two Gibbs steps from the densities:

$$f(\underline{b}_f^{(p)}, \mathbf{x}_{f,1:N}^{(p)}, \mathbf{m}_{f,1:N}^{(p)}, \boldsymbol{\delta}_{f,1:N}^{M,(p)} | \mathbf{I}, -) \quad (28)$$

$$\begin{aligned} &= f(\underline{b}_f^{(p)} | \mathbf{I}, \mathbf{x}_{f,1:N}^{(p)}, \mathbf{m}_{f,1:N}^{(p)}, \boldsymbol{\delta}_{f,1:N}^{M,(p)}, -) \\ &\quad \cdot \prod_i f(\underline{x}_{f,i}^{(p)}, \underline{m}_{f,i}^{(p)} | \boldsymbol{\delta}_{f,i}^{M,(p)}, \mathbf{I}, -) f(\boldsymbol{\delta}_{f,i}^{M,(p)} | \mathbf{I}, -) \\ &f(\mathbf{g}_{f,1:N}^{(p)}, \boldsymbol{\delta}_f^{G,(p)} | \mathbf{I}, -) \quad (29) \\ &= f(\boldsymbol{\delta}_f^{G,(p)} | \mathbf{I}, -) \prod_i f(\underline{g}_{f,i}^{(p)} | \mathbf{I}, \boldsymbol{\delta}_f^{G,(p)}, -) \end{aligned}$$

In both of the sampling steps in equations (28) and (29), we have an exact inference algorithm over multivariate-Gaussian distributed variables and Bernoulli distributed variables. This leads to faster convergence of the Markov chain and subsequently fewer burn-in samples. The conditional density for the nuisance parameters  $\boldsymbol{\eta}$  given the remainder variables can also be re-written to allow for efficient sampling. In particular, due to conditional independence we have:

$$\begin{aligned} f(\boldsymbol{\eta} | \mathbf{I}, -) &= f(\boldsymbol{\Gamma}^M | \mathbf{M}, \boldsymbol{\Delta}^M) f(\boldsymbol{\pi}^M | \mathbf{M}, \boldsymbol{\Delta}^M, \boldsymbol{\Gamma}^M) \\ &\quad \cdot f(\boldsymbol{\Gamma}^G | \mathbf{G}, \boldsymbol{\Delta}^G) f(\boldsymbol{\pi}^G | \mathbf{G}, \boldsymbol{\Delta}^G, \boldsymbol{\Gamma}^G) \\ &\quad \cdot \prod_j f(\boldsymbol{\Gamma}^B(j) | \mathbf{B}, \mathbf{C}) f(\boldsymbol{\Gamma}^X(j) | \mathbf{X}, \mathbf{C}) \end{aligned} \quad (30)$$

where  $\boldsymbol{\Gamma}$  represents the parameters related to the covariance matrices (i.e., the variance  $\sigma^2$ , correlation structure  $\boldsymbol{\Gamma}_\rho$ , and the coherence  $\rho$ ). Once again, this decomposition allows for a sampling procedure that leads to faster convergence of the Gibbs sampler. Moreover, the sampling procedures for the individual densities in equation (30) tend to require sufficient statistics that are of significantly smaller dimension and thus more desirable from a computational viewpoint. For example, sampling of the covariance matrix  $\boldsymbol{\Gamma}^M$  depends only on a  $K \times K$  sample covariance matrix. It should be noted that sampling of the covariance matrices requires additional effort in order to constrain its shape to that of equation (11). In particular, we use a Metropolis-Hastings step, which can be easily done by noting that the posterior density  $f(\boldsymbol{\Gamma}^W, \rho^W, (\sigma^2)^W | \mathbf{W})$  is proportional to an Inverse-Wishart distribution. Details are provided in Appendix B.

## VI. PERFORMANCE ANALYSIS

### A. Simulation

We first demonstrate the performance of the proposed algorithm, which we refer to as the Bayes SAR algorithm, on a simulated dataset. Images were created according to the model given in Section III with parameters given in Table VI. The low-dimensional component was divided into one of two classes ('dim' or 'bright'). Pixels were deterministically assigned to one of these classes to resemble a natural SAR image (see Figure 4). The sparse component included a randomly placed target with multiple-pixel extent. A spatiotemporally varying antenna gain filter was uniformly drawn at random on the range  $[0, 2\pi)$  for groups of pixels of size  $25 \times 25$ . Lastly, zero-mean IID noise was added with variance  $\sigma_{noise}^2$ .

The Bayes SAR model is applied to infer the low-dimensional component  $\mathbf{L}_{f,i}$  and sparse target component  $\mathbf{S}_{f,i}$  with estimates denoted  $\hat{\mathbf{L}}_{f,i}$  and  $\hat{\mathbf{S}}_{f,i}$ , respectively. Hyperparameters of the model are chosen according to the Section V. Results are given by the mean of MCMC inference with 500 burn-in iterations followed by 100 collection samples. We consider three metrics to evaluate the reconstruction errors:  $\frac{\|\mathbf{L} - \hat{\mathbf{L}}\|_2}{\|\mathbf{L}\|_2}$ ,  $\frac{\|\mathbf{S} - \hat{\mathbf{S}}\|_2}{\|\mathbf{S}\|_2}$ ,  $\frac{\|\mathbf{S} - \hat{\mathbf{S}}\|_2}{\|\mathbf{S}\|_0}$ , where the norm is taken over the vectorized quantities.

In comparison to the Bayes SAR model, results are given for state-of-the-art algorithms for Robust Principal Component Analysis (RPCA): an optimization-based approach proposed by Wright et al. [20]

TABLE VI  
PARAMETERS OF SIMULATED DATASET

Parameter	Value
Pixels in image, $P$	$P = 100 \times 100$
Number of frames per pass, $F$	$F = 1$
# of antennas, $K$	$K = 3$
# of passes, $N$	$N \in \{5, 10, 20\}$
# of target pixels/image, $N_{targets}$	$N_{targets} = 20$
Clutter of background, $\rho$	$\rho \in \{0.9, 0.99, 0.999, 0.9999\}$
Variance of targets, $\sigma_{target}^2$	$\sigma_{target}^2 = 1$
Variance of background	Either $\sigma_{dim}^2 = \sigma_{clutter}^2/100$ or $\sigma_{bright}^2 = \sigma_{clutter}^2$
Signal-to-noise-plus clutter (SCNR)	$SCNR \triangleq \frac{\sigma_{target}^2}{\sigma_{clutter}^2 + \sigma_{noise}^2}$ $\in \{0.1, 0.5, 1, 2\}$

TABLE VII

COMPARISON OF PROPOSED METHOD (BAYES SAR) TO RPCA METHODS WITH  $N = 20$ ,  $F = 1$ ,  $K = 3$ . NOTE THAT THE BAYES SAR METHOD PERFORMS ABOUT TWICE AS WELL AS EITHER OF THE RPCA METHODS FOR ALL CRITERIA. THE BAYES SAR METHOD ALSO PRODUCES A SPARSE RESULT.

(a) Bayes SAR					(b) Opt. RPCA				
SCNR	Coherence	$\frac{\ \mathbf{L} - \hat{\mathbf{L}}\ _2}{\ \mathbf{L}\ _2}$	$\frac{\ \mathbf{S} - \hat{\mathbf{S}}\ _2}{\ \mathbf{S}\ _2}$	$\frac{\ \mathbf{S} - \hat{\mathbf{S}}\ _0}{\ \mathbf{S}\ _0}$	SCNR	Coherence	$\frac{\ \mathbf{L} - \hat{\mathbf{L}}\ _2}{\ \mathbf{L}\ _2}$	$\frac{\ \mathbf{S} - \hat{\mathbf{S}}\ _2}{\ \mathbf{S}\ _2}$	$\frac{\ \mathbf{S} - \hat{\mathbf{S}}\ _0}{\ \mathbf{S}\ _0}$
10%	0.900	0.057	0.578	0.550	10%	0.900	0.111	3.175	111.026
10%	0.9999	0.045	0.419	0.367	10%	0.9999	0.113	3.237	109.716
100%	0.900	0.057	0.155	0.150	100%	0.900	0.111	1.189	109.520
100%	0.9999	0.052	0.122	0.096	100%	0.9999	0.110	1.173	108.203
200%	0.900	0.057	0.123	0.137	200%	0.900	0.112	1.058	111.120
200%	0.9999	0.056	0.114	0.092	200%	0.9999	0.110	1.035	109.583

(c) Bayes RPCA				
SCNR	Coherence	$\frac{\ \mathbf{L} - \hat{\mathbf{L}}\ _2}{\ \mathbf{L}\ _2}$	$\frac{\ \mathbf{S} - \hat{\mathbf{S}}\ _2}{\ \mathbf{S}\ _2}$	$\frac{\ \mathbf{S} - \hat{\mathbf{S}}\ _0}{\ \mathbf{S}\ _0}$
10%	0.900	0.117	0.998	3.761
10%	0.9999	0.108	0.990	3.799
100%	0.900	0.116	0.764	3.451
100%	0.9999	0.117	0.741	3.494
200%	0.900	0.125	0.706	3.665
200%	0.9999	0.129	0.692	3.720

and Candes et al. [22] and a Bayesian-based approach proposed by Ding et al. [23]<sup>2</sup>. The optimization-based approach requires a tolerance parameter which is related to the noise level, as suggested by Ding et al. [23]. We chose this parameter in order to have the smallest reconstruction errors. The Bayesian method did not require tuning parameters, except for choosing the maximum rank of  $\mathbf{L}_{f,i}$  which was set to 20.

Figure 3 compares the relative reconstruction error of the sparse (target) component,  $\frac{\|\mathbf{S} - \hat{\mathbf{S}}\|_2}{\|\mathbf{S}\|_2}$ , across all algorithms, number of passes  $N$ , coherence of antennas  $\rho$ , and SCNR. In all cases, the Bayes SAR method outperforms the RPCA algorithms with improving performance if either coherence or SCNR

<sup>2</sup>For the optimization-based approach, we used the exact\_alm\_rpca package (MATLAB) by Lin et al. [21], downloaded from <http://watt.csl.illinois.edu/perceive/matrix-rank/home.html>. For the Bayesian-based approach, we used the Bayesian robust PCA package, downloaded from [http://www.ece.duke.edu/~lihan/brpca\\_code/BRPCA.zip](http://www.ece.duke.edu/~lihan/brpca_code/BRPCA.zip).

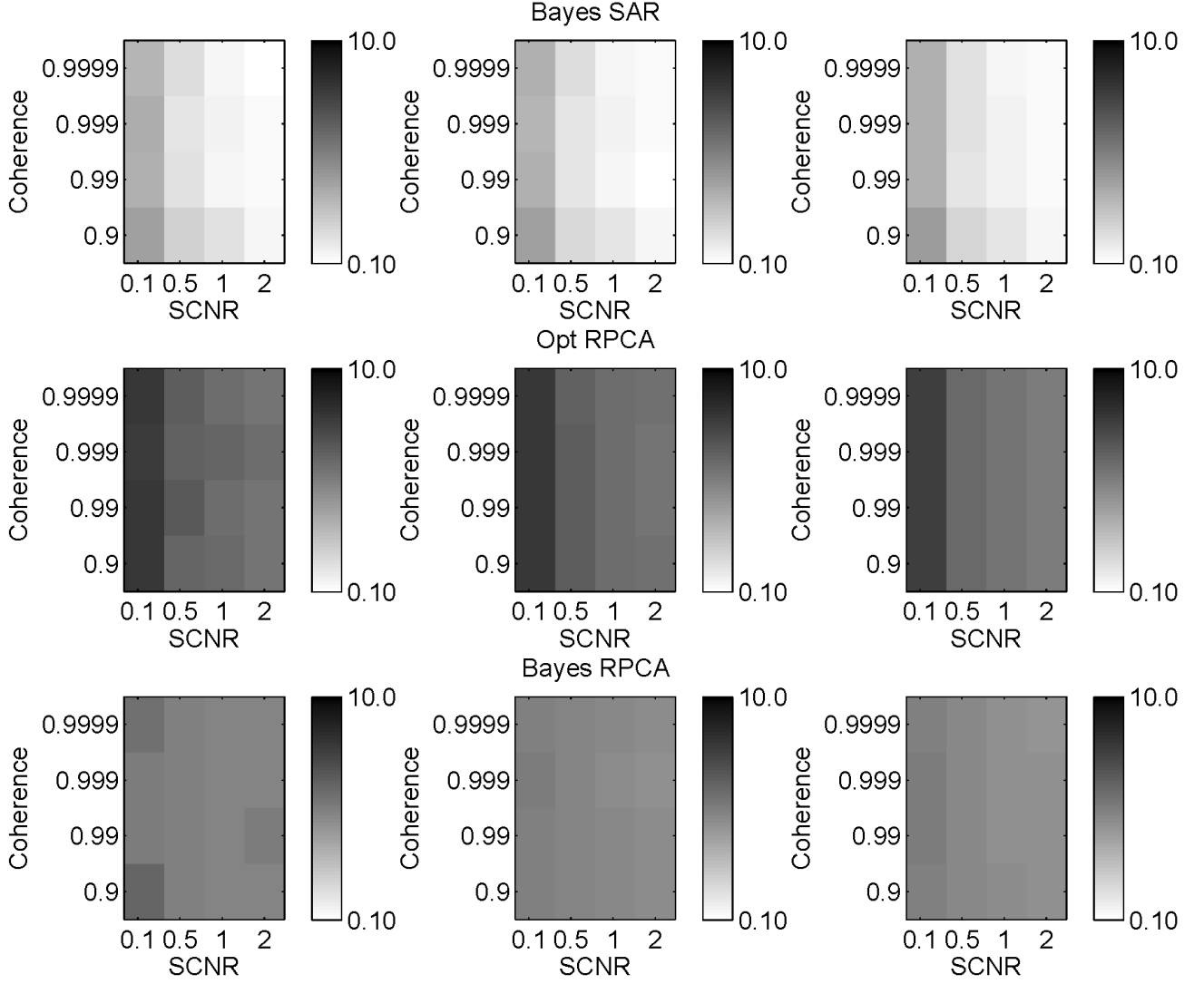


Fig. 3. This figure compares the relative reconstruction error of the target component,  $\frac{\|S - \hat{S}\|_2}{\|S\|_2}$ , as a function of algorithm, number of passes  $N$ , coherence of antennas  $\rho$ , and signal-to-clutter-plus-noise ratio (SCNR). From top-to-bottom, the rows contains the output of the Bayes SAR algorithm, the optimization-based RPCA algorithm, and the Bayes RPCA algorithm. From left-to-right, the columns show the output for  $N = 5$ ,  $N = 10$ , and  $N = 20$  passes (with  $F = 1$  frames per pass). The output is given by the median error over 20 trials on a simulated dataset. It is seen that in all cases, the Bayes SAR method outperforms the RPCA algorithms. Moreover, the Bayes SAR algorithm performs better if either coherence increases (i.e., better clutter cancellation) or the SCNR increases. On the other hand, the performance of the RPCA algorithms does not improve with increased coherence, since these algorithms do not directly model this relationship.

increases. Table VII provides additional numerical results for the case  $N = 20$ . The RPCA algorithms perform poorly in reconstructing the sparse component with relative errors near or greater than 1. This reflects the fact that (a) these algorithms miss significant sources of information, such as the correlations among antennas and among quadrature components, and (b)  $N = 20$  may be too few samples to reliably estimate the principal components in these non-parametric models. In measured SAR imagery, it might be unreasonable to expect  $N \gg 20$  passes of the radar, which suggests that these RPCA algorithms will likely perform poorly on such signals. In contrast, it is seen that the Bayes SAR method obtains low reconstruction errors for both low-dimensional and sparse components as either coherence or SCNR increase.

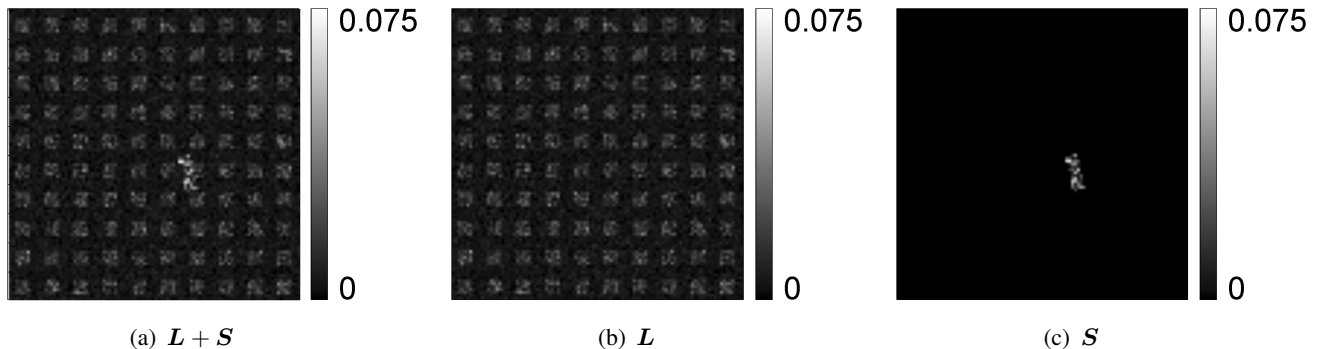


Fig. 4. This figure provides a sample image used in the simulated dataset for comparisons to RPCA methods, as well as its decomposition into low-dimensional background and sparse target components. This low SCNR image is typical of measured SAR images. Note that the target is randomly placed within the image for each of  $N$  passes. In some of these passes, the target is placed over low-amplitude clutter and can be easily detected. In other passes, the target is placed over high-amplitude clutter, which reduces the capability to detect the target.

### B. Measured data

In this section, we compare performance of the Bayes SAR approach using a set of measured data from the 2006 Gotcha SAR sensor collection. In particular, images were formed from phase histories collected over a scene of size 375m by 1200m for  $N = 3$  passes and  $K = 3$  antennas. Each image was created with a coherent processing time of 0.5 seconds with the addition of a Blackman-Harris window in the azimuth direction to reduce sidelobes. Images were created at overlapping intervals spaced 0.25 seconds apart for a total of 18 seconds. Note that the ability to take advantage of correlated images (as in this case) is one of the benefits of using the proposed model/inference algorithm.

We consider three alternative approaches in comparison to the Bayes SAR approach: (1) displaced-phase center array (DPCA) processing, (2) along-track interferometry (ATI), and (3) a mixture of DPCA/ATI. Note that all variants of ATI/DPCA depend on the chosen thresholds for phase/magnitude, respectively.

1) *Comparisons to DPCA/ATI:* We begin by comparing the output of the proposed algorithm across the entire 375m by 1200m scene. Figure 5 shows the output of the Bayes SAR algorithm and the DPCA/ATI comparisons. It is seen that there are significant performance gains by using calibrated images as shown in (c) and (f) as compared to their original versions, (b) and (e), respectively. Furthermore, the proposed approach also provides a sparse output without choosing thresholds as required by DPCA/ATI. Note that in this figure, calibration is accomplished by using the outputs  $\mathbf{H}_{f,i}$  from the Bayes SAR approach.

Figures 6 and 7 display the detection performance over two smaller scenes of size 125m by 125m as a function of magnitude and phase, respectively. For each scene, images are provided for sequential scenes separated by 0.5 seconds. Scene 1 contains strong clutter in the upper left region, while Scene 2 has relatively little clutter. It is seen that the proposed approach (2nd column) provides a sparse solution containing the targets of interest in each of the 4 images. Moreover, the 3rd column provides the estimated probability that a target occupies a given pixel, in comparison to the (0,1) output of DPCA. Although most estimated probabilities are near 1, there are a few cases where this is not the situation: in scene 2(d), a low-magnitude target is detected with low probability in the lower-right; in scene 1(b) a few target pixels from the clutter region are detected with low probability. In contrast, the performance of DPCA depends strongly on the threshold. In Scene 1, a 30 dB threshold provides a large number of false alarms. However, in Scene 2, the low-magnitude targets are missed for the 15 dB threshold, but detected at the 30 dB threshold.

Figure 7 shows the detection performance based on phase over the same 4 images. It is once again seen that the performance of the ATI/DPCA algorithms depend strongly on the thresholds, with performance that varies across thresholds from image to image. On the other hand, the proposed approach is able to detect the targets with high fidelity regardless of the scene/image and does not require tuning of thresholds for detection.



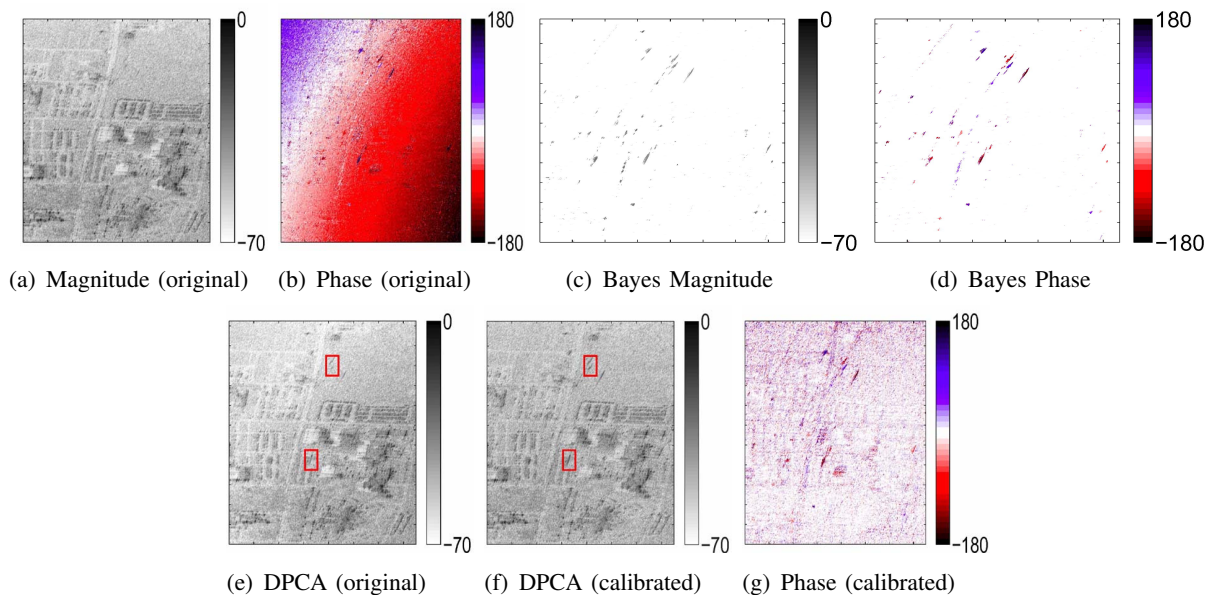


Fig. 5. This figure compares the output of the proposed algorithm as a function of magnitude and phase for a scene of size 375m by 1200m and coherent processing interval of 0.5s. The Bayes SAR algorithm takes the original SAR images in (a) and (b), estimates the nuisance parameters such as antenna miscalibrations and clutter covariances, and yields a sparse output for the target component in (c) and (d). In contrast, the DPCA and ATI algorithms are very sensitive to the nuisance parameters, which make finding detection thresholds difficult. In particular, consider the original interferometric phase image shown in (b). It can be seen that without proper calibration between antennas, there is strong spatially-varying antenna gain pattern that makes cancellation of clutter difficult. Calibration is generally not a trivial process, but to make fair comparisons to the DPCA/ATI algorithms, calibration in (f) and (g) is done by using the estimated coefficients  $\mathbf{H}_{f,i}$  from the Bayes SAR algorithm. In (e) and (f), the outputs of the DPCA algorithm are applied to the original images (all antennas) and the calibrated images (all antennas), respectively. It should be noted that even with calibration, the DPCA outputs contain a huge number of false detections in high clutter regions. Nevertheless, proper calibration enables detection of moving targets that are not easily detected without calibration, as highlighted by the red boxes. Note that the Bayes SAR algorithm provides an output that is sparse, yet does not require tuning of thresholds as required by DPCA and/or ATI.

TABLE VIII  
RADIAL VELOCITY ESTIMATION (M/S) IN 2006 GOTCHA COLLECTION DATASET

(a) Target 1				(b) Target 2			
Algorithm	Bias	MSE	No. Missed	Algorithm	Bias	MSE	No. Missed
Raw	0.64	0.94	1	Raw	0.47	0.77	6
Calibrated	0.71	1.02	0	Calibrated	0.48	0.79	0
Bayes SAR	0.03	0.10	0	Bayes SAR	0.19	0.22	0
ATI/DPCA*	-0.04	0.20	27	ATI/DPCA*	-0.07	0.43	30
ATI/DPCA**	0.10	0.20	2	ATI/DPCA**	0.23	0.28	3

2) *Target motion models*: Figure 8 shows the output of the proposed approach when prior information on the location of the targets might be available. For example, in the shown scene, targets are likely to be stopped at an intersection. The performance improvement is given for a mission scene that contains target in this high probability region. On the other hand, there are no significant performance decreases in the reference scene that does not contain targets in the intersection region. This type of processing could be extended to a tracking environment, where targets are projected to likely be in a given location within the formed SAR image as discussed in Section IV.

3) *Estimation of radial velocity*: The dataset used in this section contained a few GPS-truthed vehicles from which we can derive (a) the ‘true’ location of the target within the formed SAR image, and (b) the target’s radial velocity which is known to be proportional to the measured interferometric phase of the target pixels in an along-track system. Figure 9 shows the estimated radial velocities for two targets over

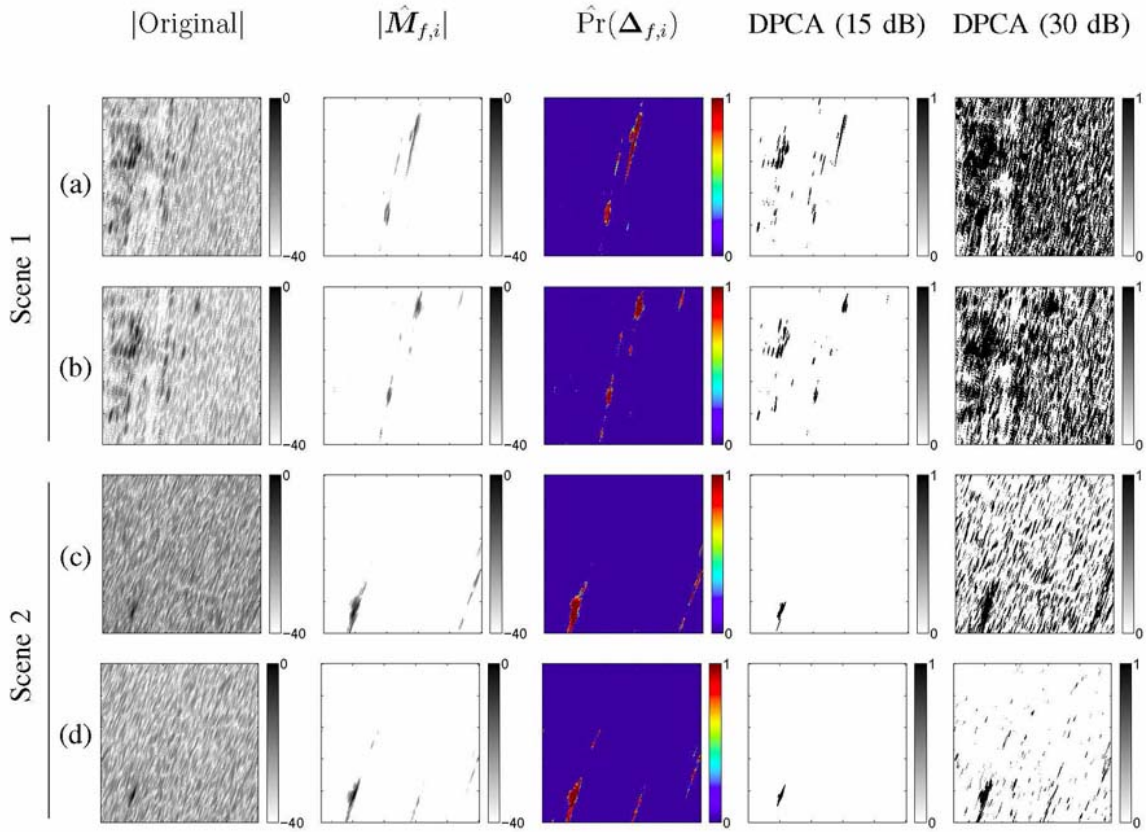


Fig. 6. This figure shows detection performance based on the magnitude of the target response with comparisons between the proposed Bayes SAR algorithm and displaced phase center array (DPCA) processing. Note that DPCA declares a detection if the relative magnitude to the brightest pixel is greater than some threshold. Results are given for two scenes of size 125m x 125m; within each scene, images were formed for two sequential 0.5 second intervals. Scene 1 contains strong clutter in the upper left region, while Scene 2 has relatively little clutter. The columns of the figure provide from left-to-right: the magnitude of the original image, the estimated target component from the proposed algorithm, the probability of the target occupying a particular pixel, the output of DPCA with a relative threshold of 15 dB, and the output of DPCA with a relative threshold of 30 dB. It is seen that DPCA has difficulty in canceling the clutter in Scene 1 with either threshold. Moreover, in Scene 2 (c-d) DPCA misses detections of the low-magnitude target in the lower right for the 15 dB threshold. In both scenes, there are many false alarms at the 30 dB threshold. On the other hand, the proposed algorithm provides a sparse solution that detects all of these targets, while simultaneously providing an estimate of the probability of detection rather than an indicator output.

18 seconds at 0.25 second increments. We compare the estimation of radial velocity from the output of the Bayes SAR algorithm, from the raw images, from the calibrated images, and from two DPCA/ATI joint algorithms with phase/magnitude thresholds of (25 deg, 15 dB) and (25 deg, 30 dB) respectively. For fair comparisons, the DPCA/ATI thresholds are applied to the calibrated imagery, though this is a non-trivial step in general. Numerical results are summarized in Table VIII. It is seen that the Bayes SAR algorithm outperforms the others in terms of MSE for both targets. Moreover, the Bayes SAR algorithm never misses a target detection in this dataset, which is not the case for the DPCA/ATI algorithms.

## VII. DISCUSSION AND FUTURE WORK

Recent work [20]–[22] has shown that it is possible to successfully decompose natural high-dimensional signals/images into low-rank and sparse components in the presence of noise, leading to the so-called robust principal component analysis algorithms. [23] introduced a Bayesian formulation of the problem that built on the success of these algorithms with the additional benefits of (a) robustness to unknown densely distributed noise with noise statistics that can be inferred from the data, (b) convergence speeds in real applications of the mean solution that are similar to those of the optimization-based procedures, and (c) characterization of the uncertainty (i.e., estimates of the posterior distribution) that could lead

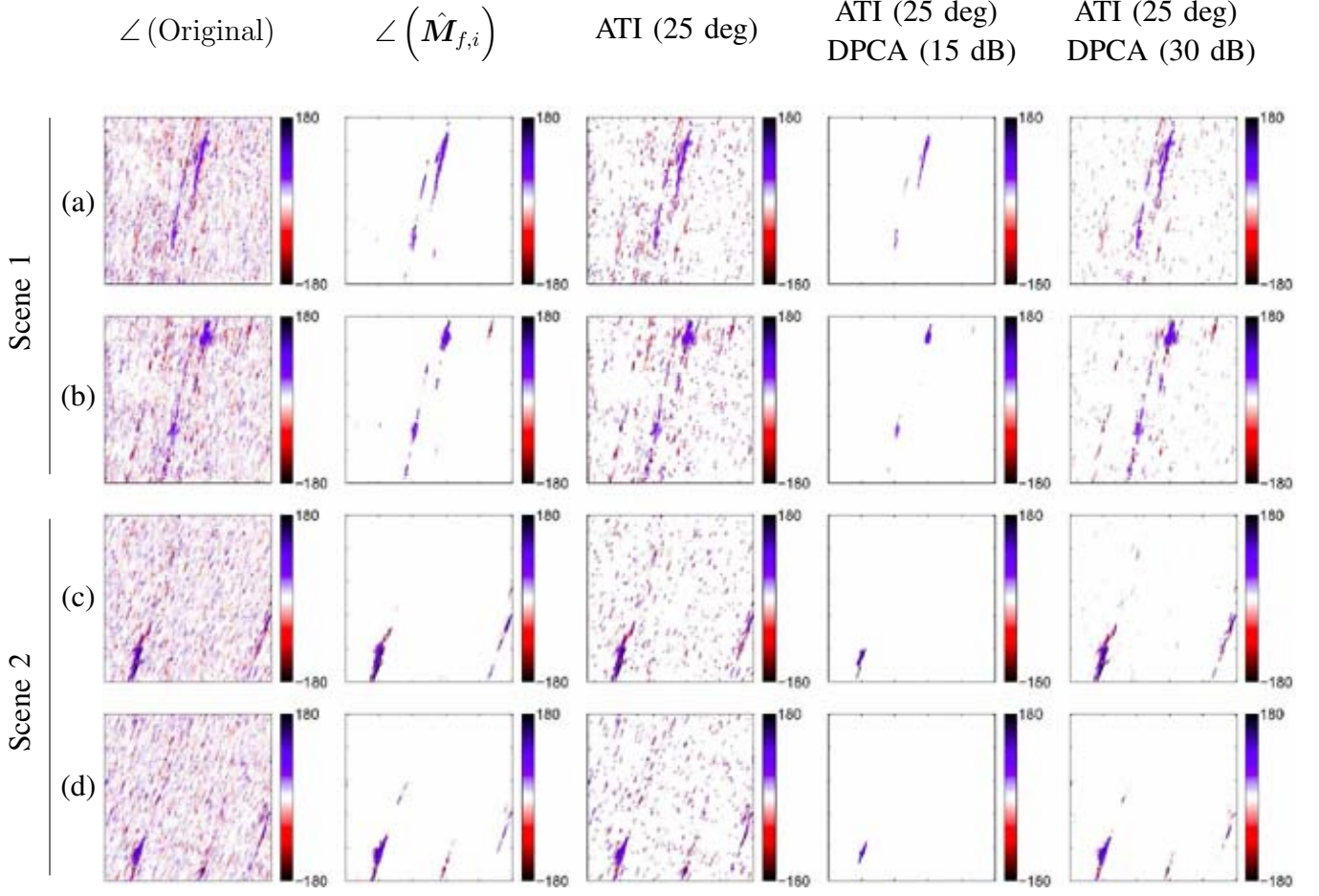


Fig. 7. This figure shows detection performance based on the phase of the target response with comparisons between the proposed algorithm, along-track interferometry (ATI) and a mixture algorithm between ATI/DPCA. Results are given for the same two scenes in Figure 6. In all cases, we show results for calibrated imagery where  $\mathbf{H}_{f,i}$  are given by the output of the Bayes SAR algorithm, though this step is not trivial. The columns of the figure provide from left-to-right: the phase of the image without thresholding, the estimated target phase component from the proposed algorithm, the output of ATI with a threshold of 25 degrees, the output of ATI/DPCA with (25 deg, 15 dB) thresholds, and the output of ATI/DPCA with (25 deg, 30 dB) thresholds. In contrast to Figure 6, the contributions from the strong clutter are not very strong, though there are still numerous false alarms in the ATI and ATI/DPCA outputs. It is seen that the ATI/DPCA combination with 15 dB magnitude threshold over-sparsifies the solution, missing targets in (b), (c), and (d). On the other hand, the ATI/DPCA combination with 30 dB magnitude threshold detects these targets, but also includes false alarms in (a) and (b).

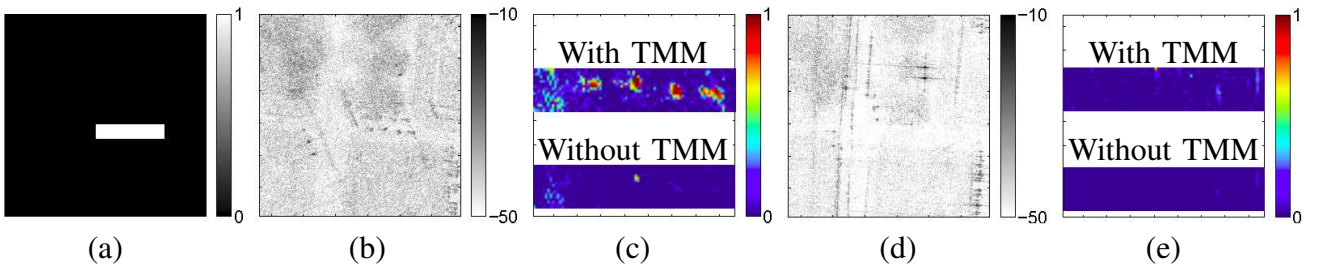


Fig. 8. This figure compares the performance of our proposed method with and without priors on target signature locations. In this scene, targets are likely to be stopped at an intersection as shown by the region in (a). A mission image containing targets is shown in (b) and a reference image without targets is shown in (d). The estimated target probabilities are shown in (c) for the mission scene where inference was done both with/without a target motion model (TMM). It can be seen that by including the prior information, we are able to detect stationary targets that cannot be detected from standard SAR moving target indication algorithms. The estimated target probabilities in the reference scene are shown in (e), showing little performance differences when prior information is included in the inference.

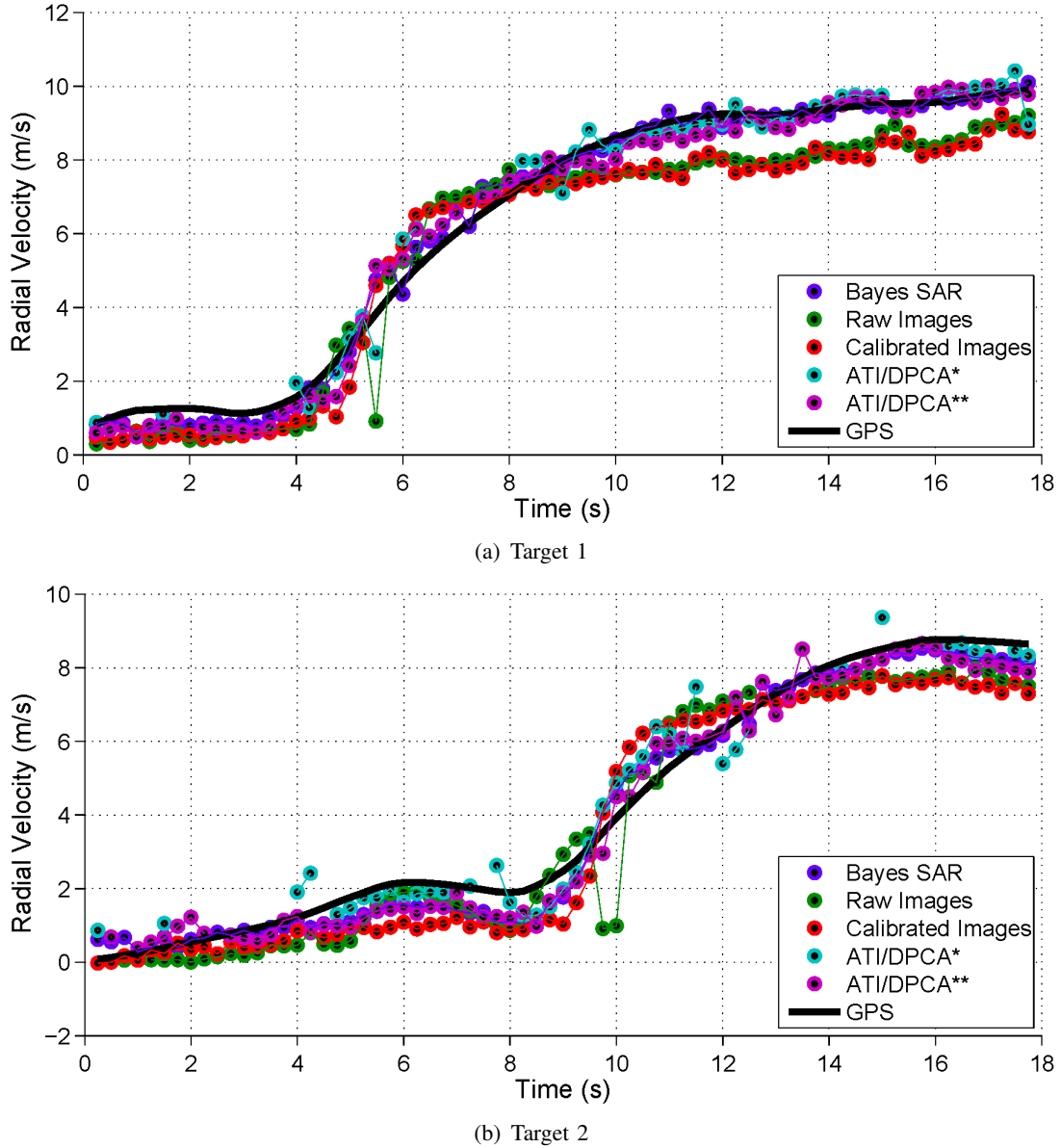


Fig. 9. This figure plots the estimated radial velocities (m/s) for two targets from measured SAR imagery over 18 seconds at 0.25 second increments. Radial velocity, which is proportional to the interferometric phase of the pixels from multiple antennas in an along-track SAR system, is estimated by computing the average phase of pixels within a region specified by the GPS-given target state (position, velocity). We compare the estimation of radial velocity from the output of the Bayes SAR algorithm, from the raw images, from the calibrated images (i.e, using the estimated calibration coefficients), and from two DPCA/ATI joint algorithms with phase/magnitude thresholds of (25 deg, 15 dB) and (25 deg, 30 dB) respectively. For best comparisons, the DPCA/ATI thresholds are applied to the calibrated imagery, though this is a non-trivial step in general. The black line provides the GPS provided radial velocities. Numerical results are summarized in Table VIII. It is seen that the Bayes SAR algorithm outperforms the others in terms of MSE for both targets. Moreover, the Bayes SAR algorithm never misses a target detection in this dataset, which is not the case for the DPCA/ATI algorithms.

to improvements in subsequent inference. Moreover, the Bayesian formulation is shown to be capable of generalization to cases where additional information is available, e.g. spatial/Markov dependencies. Future work will include the development of algorithms that exploit the use of a posterior distribution for improved performance in a signal processing task, e.g. detection, tracking or classification. In particular, we are interested in using algorithms for simultaneously detecting and estimating targets over a sparse scene with resource constraints, as well determining the fundamental performance limits of a SAR target tracking system. Furthermore, we would also like to consider other generalizations to the SAR image model, such as complex target maneuvers, multiple target classes, and explicit tracking of the target



phase.

## APPENDIX A TARGET SIGNATURE PREDICTION

In some applications, such as target tracking or sequential detection, we may have access to an estimate of the kinematic state of the target(s) of interest, such as position, velocity and acceleration. This may be useful for predicting the location of the target at sequential frames. For simplicity, consider a single target whose state  $(\mathbf{r}(\tau), \dot{\mathbf{r}}(\tau))$  is known with standard errors  $(\sigma_r, \sigma_{\dot{r}})$ , where  $\tau$  denotes the slow-time (i.e., time of the radar pulse). In standard SAR image formation, moving targets tend to appear displaced and defocused in as described in the literature by Fienup [1] and Jao [7]. Moreover, Jao shows that given the radar trajectory  $(\mathbf{q}, \dot{\mathbf{q}})$  and the target trajectory  $(\mathbf{r}, \dot{\mathbf{r}})$ , one can predict the location of the target signature within the image  $\mathbf{p}$  by solving a system of equations that equate Doppler shifts and ranges, respectively, at each pulse:

$$\frac{d}{d\tau} [\|\mathbf{p} - \mathbf{q}(\tau)\|_2 - \|\mathbf{r}(\tau) - \mathbf{q}(\tau)\|_2]_{\mathbf{p}=\mathbf{p}^*} = 0 \quad (31)$$

$$\|\mathbf{p}^* - \mathbf{q}(\tau)\|_2 = \|\mathbf{r}(\tau) - \mathbf{q}(\tau)\|_2, \quad (32)$$

which can be reduced to the simpler system of equations:

$$\dot{\mathbf{q}}(\tau) \cdot [\mathbf{p}^* - \mathbf{q}(\tau)] = [\dot{\mathbf{r}}(\tau) - \dot{\mathbf{q}}(\tau)] \cdot [\mathbf{r}(\tau) - \mathbf{q}(\tau)] \quad (33)$$

$$\|\mathbf{p}^* - \mathbf{q}(\tau)\|_2 = \|\mathbf{r}(\tau) - \mathbf{q}(\tau)\|_2 \quad (34)$$

In practice, the target state  $(\mathbf{r}, \dot{\mathbf{r}})$  is unknown or known with some uncertainty. In the latter case, we can predict the probable locations of the target signature by one of several methods, including:

- Monte Carlo estimation of the target signature locations.
- Gaussian approximation using linearization or the unscented transformation.
- Analytical approximation as proposed by Newstadt et al. [26].

### A. Notation

Following the derivation of Jao [7], we will assume the following notation:

- $\mathbf{r}(\tau) = (r_x, r_y, r_z)$  is the position of a point scatterer.
- $\mathbf{q}(\tau) = (q_x, q_y, q_z)$  is the position of the radar platform.
- $\dot{\mathbf{r}}(\tau) = (\dot{r}_x, \dot{r}_y, \dot{r}_z)$  is the velocity of a point scatterer.
- $\dot{\mathbf{q}}(\tau) = (\dot{q}_x, \dot{q}_y, \dot{q}_z)$  is the true position of the platform.
- $\mathbf{p} = (p_x, p_y, p_z)$  is a pixel location within the image.
- $\tau$  represents the slow-time (i.e., pulse of the radar sample).

### B. Deterministic solution

In the deterministic case, where  $\mathbf{r}, \dot{\mathbf{r}}, \mathbf{q}$ , and  $\dot{\mathbf{q}}$  are all known, we can find the pixel  $\mathbf{p}^*$  where the target signature will be focused at time  $\tau$  by solving equations (25) and (26). In particular, if we assume that  $z$ -coordinate is given by a function

$$p_z = h(p_x, p_y), \quad (35)$$

then we can give explicit expressions for  $(p_x, p_y, p_z)$  in some cases of  $h$ . We will focus on the simple case where  $h(p_x, p_y) = z_0$  (i.e, constant elevation), though this can be easily extended to other cases (for example, with a depth elevation map).

To solve the system of equations, let

$$\alpha(\tau) = \|\mathbf{r}(\tau) - \mathbf{q}(\tau)\|_2^2 \quad (36)$$

$$\beta(\tau) = \dot{\mathbf{q}}(\tau) \cdot \mathbf{r}(\tau) - \dot{\mathbf{r}}(\tau) \cdot (\mathbf{r}(\tau) - \mathbf{q}(\tau)) \quad (37)$$

Then we have

$$\begin{aligned}\alpha(\tau) &= \|\mathbf{p}^* - \mathbf{q}(\tau)\|_2^2 \\ &= (p_x^* - q_x(\tau))^2 + (p_y^* - q_y(\tau))^2 + (z_0 - q_z(\tau))^2\end{aligned}\quad (38)$$

and re-arranging equation (25) we have

$$\begin{aligned}\beta(\tau) &= \dot{\mathbf{q}}(\tau) \cdot \mathbf{p}^* \\ &= \dot{q}_x p_x^* + \dot{q}_y p_y^* + \dot{q}_z z_0\end{aligned}\quad (39)$$

For this derivation, assume that  $\dot{q}_x \neq 0$ <sup>3</sup>. Therefore, solving for  $p_x^*$ , we get:

$$\begin{aligned}p_x^* &= \frac{\beta(\tau) - \dot{q}_y p_y^* - \dot{q}_z z_0}{\dot{q}_x} \\ &= \left( \frac{\beta(\tau) - \dot{q}_z z_0}{\dot{q}_x} \right) + \left( -\frac{\dot{q}_y}{\dot{q}_x} \right) p_y^* \\ &= \gamma_0 + \gamma_1 p_y^*\end{aligned}\quad (40)$$

Plugging into equation (38) we get:

$$\alpha(\tau) = (\gamma_0 + \gamma_1 p_y^* - q_x(\tau))^2 + (p_y^* - q_y(\tau))^2 + (z_0 - q_z(\tau))^2 \quad (41)$$

which can be re-arranged as

$$a_y (p_y^*)^2 + b_y p_y^* + c_y (p_y^*)^2 = 0 \quad (42)$$

where

$$\begin{aligned}a_y &= (1 + \gamma_1^2) \\ b_y &= 2(\gamma_0 \gamma_1 - \gamma_1 q_x(\tau) - q_y(\tau)) \\ c_y &= (\gamma_0 - q_x(\tau))^2 + (q_y(\tau))^2 + (z_0 - q_z(\tau))^2 - \alpha(\tau) \\ &= (\gamma_0^2 + z_0^2) - 2(\gamma_0 q_x(\tau) + z_0 q_z(\tau)) + \|\mathbf{q}(\tau)\|_2^2 - \|\mathbf{r}(\tau) - \mathbf{q}(\tau)\|_2^2\end{aligned}\quad (43)$$

Then  $p_y^*$  is given by the solution of the quadratic equation:

$$p_y^* = \frac{-b_y \pm \sqrt{b_y^2 - 4a_y c_y}}{2a_y} \quad (44)$$

and  $p_x^*$  is given by equation (40). This solution suggests that the target energy will generally actually appear at two locations. However, in most cases only one of these locations will be in the formed SAR image. Thus, we generally choose the solution  $(p_x^*, p_y^*)$  that is closest to the scene center  $(0, 0)$ .

Finally, we note that equations (40) and (44) provide the pixel location containing the target energy at a single pulse time,  $\tau$ . Generally, images are formed by integrating pulses over a coherent processing interval (CPI) containing multiple times  $\tau \in [T_0, T_1]$ .

<sup>3</sup>By assumption, the radar has non-zero velocity in the  $xy$ -plane. Thus, if  $\dot{q}_x = 0$ , then this derivation should be valid if we switch the  $x$  and  $y$  indices.

### C. Uncertainty model

It is unlikely that we will have perfect information regarding the target state  $\mathbf{r}(\tau)$  and  $\dot{\mathbf{r}}(\tau)$  at all times  $\tau$ . On the other hand, there are special cases where we might have some information about these states that include

- A tracking environment, where we estimate target position and velocities with standard errors,  $\sigma_r$  and  $\sigma_{\dot{r}}$ , respectively.
- A persistently monitored scene, where we have knowledge of traffic patterns or road systems. In such a case, we have prior knowledge of likely target states.

In either case, we have a characterization of likely behavior of the target kinematic state. We can represent this knowledge in many ways that could include

- A linear kinematic model, where

$$\begin{aligned}\mathbf{r}(\tau) &= \mathbf{r}_0 + \mathbf{v}\tau + \mathbf{a}\tau^2/2 \\ \mathbf{r}_0 &\sim N(\boldsymbol{\mu}_r, \sigma_r^2 \mathbf{I}) \\ \mathbf{v} &\sim N(\boldsymbol{\mu}_v, \sigma_v^2 \mathbf{I}) \\ \mathbf{a} &\sim N(\boldsymbol{\mu}_a, \sigma_a^2 \mathbf{I})\end{aligned}\tag{45}$$

- A random kinematic model, where at each time  $\tau$

$$\begin{aligned}\mathbf{r}(\tau) &\sim N(\boldsymbol{\mu}_r(\tau), [\sigma_r(\tau)]^2 \mathbf{I}) \\ \dot{\mathbf{r}}(\tau) &\sim N(\boldsymbol{\mu}_{\dot{r}}(\tau), [\sigma_{\dot{r}}(\tau)]^2 \mathbf{I})\end{aligned}\tag{46}$$

Note that both models assume that the position and velocity vectors are Gaussian distributed. However, the first model is characterized by only 6 random variables (2 each for position, velocity, and acceleration) regardless of the number of pulses. The second model, on the other hand, assumes 4 random variables for each pulse  $\tau$ . In fact, the first model can be seen as a specialization of the second model for specific structures for the mean and variance parameters.

In this document, the choice of target kinematic model depends on the inference method which we will use to derive the distribution of the target locations. In Monte Carlo sampling, the choice of model is of relatively insignificant computational burden as compared to the generation of the Monte Carlo samples. On the other hand, in the analytical approximation methods, the choice of target kinematic model is of great importance.

The goal of this section is to provide a prediction model for the locations of targets within a SAR image given a target kinematic model. In particular, we define this model through the distribution of pixel locations:

$$f(p_x, p_y)\tag{47}$$

which is assumed to have support on  $\mathbb{R}^2$ . Generally, images are formed on a discrete grid so that we should really consider a discrete distribution. However, for simplicity we consider a continuous domain in this section.

Moreover, we have to be careful how we define the distribution of pixel locations for SAR images formed by integrating multiple radar pulses. Consider a probability distribution function (PDF) for the target location at pulse  $\tau_i$  given by  $f(p_x(\tau_i), p_y(\tau_i))$ . We define the target distribution of interest as:

$$f(p_x, p_y) = \frac{1}{T} \sum_{i=1}^T f(p_x(\tau_i), p_y(\tau_i))\tag{48}$$

This is equivalent to the distribution of the target occupying location  $(p_x, p_y)$  at any of  $T$  integrated pulses. We could consider a richer description by solving for the joint distribution on  $\{p_x(\tau_i), p_y(\tau_i)\}_{i=1}^T$ . However, this will be generally very high-dimensional and might not provide any additional benefit over the distribution given by equation (48) be useful for the purposes described in the paper.

Finally, since we will solve for the distributions  $f(p_x(\tau_i), p_y(\tau_i))$  independently for each  $\tau_i$ , we will consider the random kinematic model only. In the naive situation where the distributions of target locations are independent over time, this will provide solutions that can be approximated analytically with just a few minor assumptions.

#### D. Monte Carlo prediction

The most straightforward way to approximate the distribution in equation (48) is to use Monte Carlo sampling from the linear/random target kinematic models, followed by projection of those target states into the image domain using equations (40) and (44). A Monte Carlo representation is subsequently given by the average number of samples occupying any pixel. Note that since the Monte Carlo representation will contain discrete samples, we will end up with a discrete probability mass function (PMF) rather than a PDF.

#### E. Gaussian approximation

Rather than using a potentially high-dimensional PMF or PDF representation of the target location at time  $\tau_i$ , we could consider a Gaussian approximation that represents the probability distribution with just two parameters: the mean  $\boldsymbol{\mu}_i \in \mathbb{R}^2$  and the covariance  $\boldsymbol{\Sigma}_i \in \mathbb{R}^{2 \times 2}$  for each slow-time. Then the PDF is given by a Gaussian mixture model of form:

$$f(\mathbf{p}) \approx \frac{1}{T} \sum_{i=1}^T \phi_{\mathcal{CN}}(\mathbf{p}; \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) \quad (49)$$

where  $\phi_{\mathcal{CN}}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$  is the multivariate normal distribution PDF of  $\mathbf{x}$  with mean  $\boldsymbol{\mu}$  and covariance  $\boldsymbol{\Sigma}$ .

To find the means  $\boldsymbol{\mu}_i$  and covariances  $\boldsymbol{\Sigma}_i$ , we could consider linearization of the solution to equations (40) and (44) around a particular state  $\boldsymbol{\xi}(\tau_i) \triangleq \{\mathbf{r}(\tau_i), \dot{\mathbf{r}}(\tau_i)\}$ . This is akin to the approximation made by the Extended Kalman Filter, where the state is Gaussian but the observations are non-linear. Let us define

$$g(\boldsymbol{\xi}(\tau_i)) = \begin{bmatrix} p_x^*(\boldsymbol{\xi}(\tau_i)) \\ p_y^*(\boldsymbol{\xi}(\tau_i)) \end{bmatrix} \quad (50)$$

where  $p_x^*$  and  $p_y^*$  are given by equations (40) and (44). Then the first-order linearization around the mean  $\boldsymbol{\mu}_\xi$  is given by:

$$G(\boldsymbol{\xi}) = g(\boldsymbol{\mu}_\xi) + \left( \nabla g(\boldsymbol{\xi})|_{\boldsymbol{\xi}=\boldsymbol{\mu}_\xi} \right)^T (\boldsymbol{\xi} - \boldsymbol{\mu}_\xi) \quad (51)$$

Since this is a linear function of a Gaussian distributed vector in  $\boldsymbol{\xi}(\tau_i)$ , we know that the pixels  $\mathbf{p}(\tau_i) \approx G(\boldsymbol{\xi}(\tau_i))$  are distributed as

$$\mathbf{p}(\tau_i) \sim N \left( g(\boldsymbol{\mu}_\xi(\tau_i)), \left( \nabla g(\boldsymbol{\xi})|_{\boldsymbol{\xi}=\boldsymbol{\mu}_\xi(\tau_i)} \right)^T \boldsymbol{\Sigma}_\xi(\tau_i) \left( \nabla g(\boldsymbol{\xi})|_{\boldsymbol{\xi}=\boldsymbol{\mu}_\xi(\tau_i)} \right) \right), \quad (52)$$

where  $\boldsymbol{\Sigma}_\xi(\tau_i)$  is given by

$$\boldsymbol{\Sigma}_\xi(\tau_i) = \begin{bmatrix} \boldsymbol{\Sigma}_r(\tau_i) & 0 \\ 0 & \boldsymbol{\Sigma}_{\dot{r}}(\tau_i) \end{bmatrix} \quad (53)$$



### F. Analytical approximation

It is also possible to get a closer approximation than the linearization example provided above by doing some analytical derivations. In particular, let us assume that the radar platform moves in the  $x$ -direction so that  $\dot{q}_y = \dot{q}_z = 0$  and  $|\dot{q}_x| > 0$ . In the general case, this derivation would hold for a transformed set of coordinates  $(p'_x, p'_y)$ , though we won't go into that derivation here. In the former case, equations (40) and (44) reduce to:

$$p_x^* = \frac{\beta(\tau)}{\dot{q}_x} = \frac{\dot{q}_x r_x - (r_x - q_x)v_x - (r_y - q_y)v_y}{\dot{q}_x} \quad (54)$$

$$p_y^* = q_y \pm \sqrt{\alpha(\tau) - (p_x - q_x)^2 - (z_0 - q_z)^2} \quad (55)$$

From equation (54), we see that

$$f(p_x^*|r_x, r_y) \sim N(\mu, \sigma^2) \quad (56)$$

where

$$\mu = \frac{1}{\dot{q}_x} ([\dot{q}_x - \mu_{vx}]r_x - \mu_{vy}r_y + q_x\mu_{vx} + q_y\mu_{vy}) \quad (57)$$

$$\sigma^2 = \frac{(r_x - q_x)^2 + (r_y - q_y)^2}{(\dot{q}_x)^2} \sigma_v^2 \quad (58)$$

where we have assumed that  $\Sigma_{\dot{r}} = \sigma_v^2 \mathbf{I}$ . Note that  $\sigma^2$  is a function of the position  $\mathbf{r}$ . However, since  $\|q\| \gg \|r\|$  in general, we make a zero-th order approximation here so that

$$\sigma^2 \approx \sigma_0^2 \triangleq \frac{(\mu_{rx} - q_x)^2 + (\mu_{ry} - q_y)^2}{(\dot{q}_x)^2} \sigma_v^2 \quad (59)$$

In this case, we see that we can find

$$\begin{aligned} f(p_x) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(p_x|r_x, r_y) f(r_x, r_y) dr_x dr_y \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \phi_{\mathcal{CN}}(p_x; \mu(r_x, r_y), \sigma_0^2) \phi_{\mathcal{CN}}(r_x; \mu_{rx}, \sigma_r^2) \phi_{\mathcal{CN}}(r_y; \mu_{ry}, \sigma_r^2) dr_x dr_y \end{aligned} \quad (60)$$

where we have assumed that  $\Sigma_r = \sigma_r^2 \mathbf{I}$ . Since  $\mu(r_x, r_y)$  is linear in  $r_x$  and  $r_y$  from equation (57), we can analytically solve this integral to see that

$$p_x^* \sim N(\mu_{px}, \sigma_{px}^2) \quad (61)$$

where

$$\mu_{px} = \frac{1}{\dot{q}_x} (q_x\mu_{vx} + q_y\mu_{vy} + \dot{q}_x\mu_{rx} - \mu_{vx}\mu_{rx} - \mu_{vy}\mu_{ry}) \quad (62)$$

$$\sigma_{px}^2 = \frac{\sigma_r^2}{(\dot{q}_x)^2} [(\dot{q}_x - \mu_{vx})^2 + \mu_{vy}^2] + \sigma_0^2 \quad (63)$$

Since  $p_y^*$  in equation (55) is non-linear, we make one more assumption with a first-order linearization around  $p_y^*(\boldsymbol{\mu}_r)$ . Note that given  $p_x^*$ , equation (55) only depends on the target state through  $\mathbf{r}$  (and not on the velocity  $\dot{\mathbf{r}}$ ). Define  $s(\mathbf{r}, p_x)$  to be the value of equation (55) given state  $\mathbf{r}$  and pixel location  $p_x$ . Then we approximate  $p_y^*$  as

$$p_y^*(\mathbf{r})|p_x^* \approx s(\boldsymbol{\mu}_r, p_x^*) + \left( \nabla s(\boldsymbol{\mu}_r, p_x^*)|_{\mathbf{r}=\boldsymbol{\mu}_r} \right)^T (\mathbf{r} - \boldsymbol{\mu}_r) \quad (64)$$

Finally, we note that given  $p_x, p_y$  is Gaussian distributed with distribution:

$$p_y|p_x \sim N(\mu_{py}, \sigma_{py}^2) \quad (65)$$

$$\mu_{py} = s(\boldsymbol{\mu}_r, p_x^*) \quad (66)$$

$$\sigma_{py}^2 = \sigma_r^2 \left( \nabla s(\boldsymbol{\mu}_r, p_x^*)|_{\mathbf{r}=\boldsymbol{\mu}_r} \right)^T \left( \nabla s(\boldsymbol{\mu}_r, p_x^*)|_{\mathbf{r}=\boldsymbol{\mu}_r} \right) \quad (67)$$

Note that both  $p_x^*$  and  $p_y^* | p_x^*$  have Gaussian distributions that can be described by a mean and covariance term. In contrast to Section A-E, the distribution is not jointly Gaussian because the mean and covariance of  $p_y^* | p_x^*$  depend on  $p_x^*$ . Nevertheless, one can easily evaluate this PDF at any pixel  $(p_x, p_y)$ . Over short CPIs, both approximations will probably lead to similar results.

## APPENDIX B INFERENCE DETAILS

In the hierarchical model proposed in Section III, the distribution of hyper-parameters at the base layer are generally chosen to be conjugate to the distributions at the next layer. This allows for efficient approximation methods for the posterior distribution in the sense that we can sample exactly from these distributions. In particular, we use a Markov Chain Monte Carlo (MCMC) algorithm in the form of a Gibbs sampler to iteratively estimate the full joint posterior. In MCMC, this distribution is approximated by drawing samples iteratively from the conditional distribution of each (random) model variable given the most recent estimate of the rest of the variables (which we denote by  $-$ ). Let  $\Theta = \{\mathbf{B}, \mathbf{X}, \mathbf{G}, \mathbf{M}, \Delta^G, \Delta^M, \mathbf{H}, \mathbf{C}, \boldsymbol{\eta}\}$  represent a current estimate of all of the model variables where  $\boldsymbol{\eta}$  represents the set of all hyper-parameters. Given measurements  $\mathbf{I}$ , the inference algorithm is given in Figure 2. Note that MCMC algorithms require a burn-in period after the Markov chain has become stable, where the duration of burn-in period depends on the problem. After this point, we collect  $N_{samples}$  samples that represent the full joint distribution.

The sampling details are provided for each of the steps in Figure 2 individually.

### A. Basic Decomposition

Given the parameters in Tables IV and V, we arrive at one of the primary benefits of using Gibbs Sampling for inference: namely that we can independently sample across pixels and frames. In distributional form, we have

$$\begin{aligned} & f(\mathbf{B}, \mathbf{X}, \mathbf{G}, \mathbf{M}, \Delta^G, \Delta^M | \mathbf{I}, -) \\ &= \prod_{p,f} f(\underline{b}_f^{(p)}, \mathbf{x}_{f,1:N}^{(p)}, \mathbf{g}_{f,1:N}^{(p)}, \mathbf{m}_{f,1:N}^{(p)}, \delta_f^{G,(p)}, \delta_{f,1:N}^{M,(p)} | \mathbf{I}, -) \end{aligned} \quad (68)$$

The conditional independence among pixels and frames given the nuisance parameters allows us to easily parallelize the sampling procedure over the largest dimensions of the state. Moreover, we can extend the parallelization to sampling independently over passes by separating the sampling of equation (68) into two Gibbs steps from the densities:

$$\begin{aligned} & f(\underline{b}_f^{(p)}, \mathbf{x}_{f,1:N}^{(p)}, \mathbf{m}_{f,1:N}^{(p)}, \delta_{f,1:N}^{M,(p)} | \mathbf{I}, -) \\ &= f(\underline{b}_f^{(p)} | \mathbf{I}, \mathbf{x}_{f,1:N}^{(p)}, \mathbf{m}_{f,1:N}^{(p)}, \delta_{f,1:N}^{M,(p)}, -) \\ &\quad \cdot \prod_i f(\underline{x}_{f,i}^{(p)}, \underline{m}_{f,i}^{(p)} | \delta_{f,i}^{M,(p)}, \mathbf{I}, -) f(\delta_{f,i}^{M,(p)} | \mathbf{I}, -) \end{aligned} \quad (69)$$

$$\begin{aligned} & f(\mathbf{g}_{f,1:N}^{(p)}, \delta_f^{G,(p)} | \mathbf{I}, -) \\ &= f(\delta_f^{G,(p)} | \mathbf{I}, -) \prod_i f(\underline{g}_{f,i}^{(p)} | \mathbf{I}, \delta_f^{G,(p)}, -) \end{aligned} \quad (70)$$

It should be noted that each of these distributions have explicit forms as either multivariate Gaussian or Bernoulli distributed. For example consider the distribution

$$f(\underline{x}_{f,i}^{(p)}, \underline{m}_{f,i}^{(p)} | \delta_{f,i}^{M,(p)}, \mathbf{I}, -) = f(\underline{x}_{f,i}^{(p)}, \underline{m}_{f,i}^{(p)} | \delta_{f,i}^{M,(p)}, \underline{r}_{f,i}^{(p)}, -) \quad (71)$$

where we define

$$\underline{r}_{f,i}^{(p)} \triangleq \left( \underline{i}_{f,i}^{(p)} / \underline{h}_{f,i}^{(p)} \right) - \delta_f^{G,(p)} \underline{g}_{f,i}^{(p)}, \quad (72)$$

where  $./$  is the point-wise division operator. In particular, we know that

$$\underline{r}_{f,i}^{(p)} = \underline{b}_f^{(p)} + \underline{x}_{f,i}^{(p)} + \delta_{f,i}^{M,(p)} \underline{m}_{f,i}^{(p)} + \underline{v}_{f,i}^{(p)}, \quad (73)$$

Thus, given  $\delta_{f,i}^{M,(p)}$ , we have  $\underline{r}_{f,i}^{(p)}$  as a linear combination of Gaussian random variables. Thus, by standard conditional distributions of a Gaussian random vector, we know that the distribution of  $\underline{b}_f^{(p)}$  and  $\underline{x}_{f,i}^{(p)}$  must also be Gaussian. Table IX gives the means and covariances of the Gaussian distributions in equations (69) and (70). Moreover, the distributions of the indicator functions are easily found by Bayes rule. For example, define the quantity

$$z_d = f(\underline{i}_{f,i}^{(p)} | \delta_{f,i}^{M,(p)} = d, \underline{h}_{f,i}^{(p)}, \delta_f^{G,(p)}, \underline{g}_{f,i}^{(p)}, \underline{b}_f^{(p)}). \quad (74)$$

Then it is easily seen that for the target indicators we have by Bayes rule

$$\begin{aligned} f(\delta_{f,i}^{M,(p)} = 1 | \mathbf{I}, -) &= \frac{z_1 f(\delta_{f,i}^{M,(p)} = 1)}{\sum_{d=(0,1)} z_d f(\delta_{f,i}^{M,(p)} = d)} \\ &= \left[ 1 + \frac{1 - \pi_{f,i}^M(p)}{\pi_{f,i}^M(p)} \frac{z_0}{z_1} \right]^{-1} \end{aligned} \quad (75)$$

Note that  $z_d$  is just an evaluation of the Normal PDF so that it can be simply calculated. Table X provides explicit values of  $z_d$  for both the target and glint indicators, where  $\phi_{\mathcal{CN}}(x; \boldsymbol{\mu}, \boldsymbol{\Sigma})$  is the multivariate normal PDF with mean  $\boldsymbol{\mu}$  and covariance  $\boldsymbol{\Sigma}$ .

### B. Calibration coefficients

For this thesis, we assumed that pixels within a subset  $Z_g \subset \{1, 2, \dots, P\}$  share the same calibration constant so that

$$h_{k,f,i}^{(p)} = z_{k,f,i}(g), \quad \forall p \in Z_g. \quad (76)$$

with  $z_{k,f,i}(g) \sim \mathcal{CN}(1, (\sigma^H)^2)$ . In our formulation (and dropping the  $(k, f, i)$  indices for simplicity) we have measurements of the form

$$i^{(p)} = z(g)(l^{(p)} + s^{(p)} + v^{(p)}), \quad \forall p \in Z_g. \quad (77)$$

Define  $y^{(p)} = l^{(p)} + s^{(p)}$  which is a known quantity in our Gibbs sampling inference step. Moreover, we assume that  $|y^{(p)}| \gg |v^{(p)}|$  so that for any  $p \in Z_g$  we have

$$\begin{aligned} i^{(p)} &= z(g)(y^{(p)} + v^{(p)}) \\ &\approx z(g)y^{(p)} + E[z(g)]v^{(p)} \\ &= z(g)y^{(p)} + v^{(p)} \end{aligned} \quad (78)$$

TABLE IX  
GAUSSIAN DISTRIBUTION PARAMETERS FOR DISTRIBUTIONS OF BASE LAYER PARAMETERS IN SAR IMAGE MODEL EQUATIONS (69) AND (70)

Component	Variable	Mean	Covariance	Parameters
Background	$\underline{b}_f^{(p)}$	$\Lambda_{\underline{r}_f^{(p)}}$	$(\mathbf{I}_{K \times K} - \Lambda) \mathbf{\Gamma}^B(p)$	$\underline{r}_f^{(p)} = \sum_{i=1}^N (\underline{b}_{f,i}^{(p)} / \underline{h}_{f,i}^{(p)} - \underline{x}_{f,i}^{(p)} - \delta_f^{G,(p)} \underline{g}_{f,i}^{(p)} - \delta_{f,i}^{M,(p)} \underline{m}_{f,i}^{(p)})$ $\Lambda = \mathbf{\Gamma}^B(p) (\mathbf{\Gamma}^B(p) + \mathbf{\Gamma}^V)^{-1}$
Speckle, Target	$\begin{bmatrix} \underline{x}_{f,i}^{(p)} \\ \underline{m}_{f,i}^{(p)} \end{bmatrix}$	$\Lambda_{\underline{r}_{f,i}^{(p)}}$	$\Sigma_{11} - \Lambda \Sigma_{12}$	$\Sigma_{11} = \begin{bmatrix} \mathbf{\Gamma}^X(p) & 0 \\ 0 & \delta_{f,i}^{M,(p)} \mathbf{\Gamma}^M \end{bmatrix}, \Sigma_{12} = \begin{bmatrix} \mathbf{\Gamma}^X(p) \\ \delta_{f,i}^{M,(p)} \mathbf{\Gamma}^M \end{bmatrix}$ $\underline{r}_{f,i}^{(p)} = \underline{b}_{f,i}^{(p)} / \underline{h}_{f,i}^{(p)} - \underline{b}_f^{(p)} - \delta_f^{G,(p)} \underline{g}_{f,i}^{(p)}$ $\Lambda = \Sigma_{12}^T (\mathbf{\Gamma}^X(p) + \delta_{f,i}^{M,(p)} \mathbf{\Gamma}^M + \mathbf{\Gamma}^V)^{-1}$
Glints	$\underline{g}_{f,i}^{(p)}$	$\Lambda_{\underline{r}_{f,i}^{(p)}}$	$(\mathbf{I}_{K \times K} - \Lambda) \mathbf{\Gamma}^G$	$\underline{r}_{f,i}^{(p)} = \underline{b}_{f,i}^{(p)} / \underline{h}_{f,i}^{(p)} - \underline{x}_{f,i}^{(p)} - \delta_{f,i}^{M,(p)} \underline{m}_{f,i}^{(p)} - \underline{b}_f^{(p)}$ $\Lambda = \delta_f^{G,(p)} \mathbf{\Gamma}^G (\delta_f^{G,(p)} \mathbf{\Gamma}^G + \mathbf{\Gamma}^V)^{-1}$

TABLE X  
BERNOULLI DISTRIBUTION PARAMETERS FOR DISTRIBUTIONS OF INDICATOR VARIABLES IN EQUATIONS (69) AND (70)

Component	Variable	$f(\delta = 1   -)$	$z_d$	$\underline{r}^{(p)}$
Glints	$\delta_f^{G,(p)}$	$\left[ 1 + \frac{1 - \pi_f^G(p)}{\pi_f^G(p)} \frac{z_0}{z_1} \right]^{-1}$	$\phi_{\mathcal{CN}}(\underline{r}^{(p)} / N; \mathbf{0}, d\mathbf{\Gamma}^G + \mathbf{\Gamma}^V)$	$\sum_{i=1}^N \left( \frac{i_{f,i}^{(p)}}{\underline{i}_{f,i}^{(p)}} / \frac{h_{f,i}^{(p)}}{\underline{h}_{f,i}^{(p)}} - \underline{b}_f^{(p)} - \underline{x}_{f,i}^{(p)} - \delta_{f,i}^{M,(p)} \underline{m}_{f,i}^{(p)} \right)$
Targets	$\delta_{f,i}^{M,(p)}$	$\left[ 1 + \frac{1 - \pi_{f,i}^M(p)}{\pi_{f,i}^M(p)} \frac{z_0}{z_1} \right]^{-1}$	$\phi_{\mathcal{CN}}(\underline{r}^{(p)}; \mathbf{0}, d\mathbf{\Gamma}^M + \mathbf{\Gamma}^V)$	$\frac{i_{f,i}^{(p)}}{\underline{i}_{f,i}^{(p)}} / \frac{h_{f,i}^{(p)}}{\underline{h}_{f,i}^{(p)}} - \underline{b}_f^{(p)} - \underline{x}_{f,i}^{(p)} - \delta_f^{G,(p)} \underline{g}_{f,i}^{(p)}$

Note that given  $y^{(p)}$  (as in the Gibbs sampling step), we have the situation where  $i^{(p)}$ ,  $z(g)$ , and  $v^{(p)}$  are all Gaussian distributed random variables. Thus, the conditional distribution of  $z(g)$  is also Gaussian with:

$$z(g)|y \sim \mathcal{CN}(\mu_z(g), \sigma_z^2(g)) \quad (79)$$

$$\mu_z(g) = 1 + \left[ \frac{(\sigma^H)^2}{(\sigma^V)^2 + y^H y (\sigma^H)^2} \right] y^H (i - y) \quad (80)$$

$$\sigma_z^2(g) = \frac{(\sigma^H)^2 (\sigma^V)^2}{(\sigma^V)^2 + y^H y (\sigma^H)^2} \quad (81)$$

where  $y = \{y^{(p)}\}_{p \in Z_g}$  and  $i = \{i^{(p)}\}_{p \in Z_g}$ . Note that when  $(\sigma^H)^2$  is large, then maximum likelihood inference in this case yields the least-squares solution for  $z(g)$ .

### C. Object class assignment

In this model, we assume that each pixel can be assigned to one of  $J$  possible classes. We assume that the number  $J$  is known a priori and do not consider the details involved in the merging or splitting of object classes here. More detailed models (such as the so-called Indian Buffet processes) can also estimate the number of classes directly from the data.

In this model, inference on class assignment is straightforward given the distributions (i.e., covariance matrices) for each class. Define the matrices

$$\mathbf{b}^{(p)} = \begin{bmatrix} \underline{b}_1^{(p)} \\ \underline{b}_2^{(p)} \\ \vdots \\ \underline{b}_F^{(p)} \end{bmatrix} \in \mathbb{C}^{F \times K} \quad \mathbf{x}^{(p)} = \begin{bmatrix} \underline{x}_{1,1}^{(p)} \\ \underline{x}_{1,2}^{(p)} \\ \vdots \\ \underline{x}_{F,N}^{(p)} \end{bmatrix} \in \mathbb{C}^{FN \times K} \quad (82)$$

Then the probability that pixel  $p$  belongs to class  $j$  is given by:

$$w_j^{(p)} \triangleq \Pr(\text{pixel } p \text{ has class } j) = \exp\{T_B + T_X + q_j\} \quad (83)$$

where  $q_j$  is the prior probability of class  $j$  and

$$T_B = -\text{trace}([\mathbf{\Gamma}^B(j)]^{-1} \mathbf{b}^{(p)} (\mathbf{b}^{(p)})^H) - \frac{F}{2} \log |\mathbf{\Gamma}^B(j)| - KF \log(2\pi) \quad (84)$$

$$T_X = -\text{trace}([\mathbf{\Gamma}^X(j)]^{-1} \mathbf{x}^{(p)} (\mathbf{x}^{(p)})^H) - \frac{FN}{2} \log |\mathbf{\Gamma}^X(j)| - KFN \log(2\pi) \quad (85)$$

Then the class assignment to pixel  $p$  is the single location in  $\underline{c}^{(p)}$  with value equal to one, where

$$\underline{c}^{(p)} \sim \text{Multinomial}(1; \underline{w}^{(p)}) \quad (86)$$

Note that we can improve upon this model by allowing the probabilities for pixel  $p$  to vary spatially (i.e., pixels are likely to share the same class with neighboring pixels). One simple way to include this information is to let

$$\underline{m}^{(p)} = \underline{w}^{(p)} * \underline{g}_{HMM}^{(p)} \quad (87)$$

where  $\underline{g}_{HMM}^{(p)}$  is some filter for averaging nearby pixels and  $*$  is the convolution operator (assumed to be supported on the same set of pixels as  $\underline{w}^{(p)}$ ). Then we draw

$$\underline{c}^{(p)} \sim \text{Multinomial}(1; \underline{m}^{(p)}) \quad (88)$$

### D. Hyper-parameters

In this model, we have three types of hyper-parameters that need to be estimated: covariance matrices (or variances), indicator probabilities, and object class probabilities. In all cases, the distribution of these parameters depend on test statistics of much smaller dimension than  $P$ .

1) *Covariance matrix inference:* We model the covariance matrices for the Normal distributions in two ways: (1) for the stationary components (background, speckle, and glints), we model the covariance matrix as a random variable; and (2) for the other components (targets, additive noise, calibration coefficients), we assume independence among the antennas. In particular, consider a random vector of  $K$  elements,  $\underline{w}$ , with

$$\underline{w} \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{\Gamma}_\rho) \quad (89)$$

$$\sigma^2 \sim \text{InvGamma}(a_\sigma, b_\sigma) \quad (90)$$

Then in the stationary case, we have

$$\mathbf{\Gamma}_\rho \sim \text{InvWishart}(a_\Gamma((1 - \rho)\mathbf{I}_{K \times K} + \rho \mathbf{1}_K \mathbf{1}_K^T), \nu_\Gamma) \quad (91)$$

$$\rho \sim \text{Beta}(a_\rho, b_\rho) \quad (92)$$

and in the independent case, we have

$$\mathbf{\Gamma}_\rho = \mathbf{I}_{K \times K} \quad (93)$$

First consider the case where  $\mathbf{\Gamma}_\rho$  is a random variable. Assume that we have  $n$  independent samples of  $\underline{w}$ , which we refer to as  $\mathbf{W} = \text{vec}\{\underline{w}\}$ . Then, we consider a Gibbs sampling procedure:

$$\text{Sample} \sim f(\sigma^2 | \mathbf{W}, \mathbf{\Gamma}_\rho, \rho) \quad (94)$$

$$\text{Sample} \sim f(\mathbf{\Gamma}_\rho, \rho | \mathbf{W}, \sigma^2) \quad (95)$$

Let  $\tau = 1/\sigma^2 \sim \text{Gamma}(a_\sigma, b_\sigma)$ . Then

$$\begin{aligned} f(\tau | \mathbf{W}, \mathbf{\Gamma}_\rho, \rho) &\propto f(\mathbf{W} | \tau, \mathbf{\Gamma}_\rho) f(\tau) \\ &\propto \left[ \tau^{n/2} \exp \left\{ -\frac{\tau}{2} \text{trace}(\mathbf{\Gamma}_\rho^{-1} \mathbf{W} \mathbf{W}^H) \right\} \right] \left[ \frac{(b_\sigma)^{a_\sigma}}{\Gamma(a_\sigma)} \tau^{a_\sigma - 1} \exp \{-b_\sigma \tau\} \right] \\ &\propto \frac{(b')^{a'}}{\Gamma(a')} \tau^{a' - 1} \exp \{-b' \tau\} \end{aligned} \quad (96)$$

where

$$\begin{aligned} a' &= a_\sigma + \frac{n}{2} \\ b' &= b_\sigma + \frac{\text{trace}(\mathbf{\Gamma}_\rho^{-1} \mathbf{W} \mathbf{W}^H)}{2} \end{aligned} \quad (97)$$

This demonstrates that in this situation,  $\sigma^2$  has an Inverse-Gamma distribution with parameters  $a'$  and  $b'$ . Note that in the case where  $\mathbf{\Gamma}_\rho = \mathbf{I}_{K \times K}$ , then the posterior parameters are given by

$$a' = a_\sigma + \frac{n}{2} \quad (98)$$

$$b' = b_\sigma + \frac{\text{trace}(\mathbf{W} \mathbf{W}^H)}{2} \quad (99)$$

Thus, in the Gibbs sampling procedure, the variance parameter  $\sigma^2$  is Inverse-Gamma distributed whether or not  $\mathbf{\Gamma}_\rho$  is modeled as a random variable. Table XI provides the posterior Inverse Gamma distribution parameters for the variance parameters in our model, where  $\text{vec}\{\cdot\}$  refers to the vectorization operator.

For the background, speckle, and glint components, we also need to sample the coherence parameter  $\rho$  and correlation matrix  $\mathbf{\Gamma}_\rho$ . Let  $\tilde{\mathbf{W}} = \mathbf{W}/\sigma$  be our observed measurements given  $\sigma^2$  as given by equation (95). Define  $\tilde{\boldsymbol{\mu}} = \boldsymbol{\mu}/\sigma$ . Then we have

$$\begin{aligned} \tilde{\mathbf{W}} | (\mathbf{\Gamma}_\rho, \sigma^2, \rho) &\sim \mathcal{CN}(\tilde{\boldsymbol{\mu}}, \mathbf{\Gamma}_\rho) \\ \mathbf{\Gamma}_\rho &\sim \text{InvWishart}([\rho \mathbf{1}_K \mathbf{1}_K^T + (1 - \rho) \mathbf{I}_{K \times K}] (\nu - K - 1), \nu) \\ \rho &\sim \text{Beta}(a_\rho, b_\rho) \end{aligned} \quad (100)$$

TABLE XI  
INVERSE GAMMA DISTRIBUTION PARAMETERS FOR DISTRIBUTIONS OF VARIANCES AND COVARIANCE MATRIX ESTIMATES

Component	Variable	$a'$	$b'$	Other parameters
Background	$(\sigma^B(j))^2$	$a_\sigma + \frac{M}{2}$	$\frac{\text{trace}(\mathbf{\Gamma}_\rho^{-1} \mathbf{W} \mathbf{W}^H)}{2}$	$M = F Q_j , \mathbf{W} = \text{vec} \left\{ \underline{b}_f^{(p)} \right\}_{f,p \in Q_j}$
Speckle	$(\sigma^X(j))^2$	$a_\sigma + \frac{M}{2}$	$\frac{\text{trace}(\mathbf{\Gamma}_\rho^{-1} \mathbf{W} \mathbf{W}^H)}{2}$	$M = NF Q_j , \mathbf{W} = \text{vec} \left\{ \underline{x}_{f,i}^{(p)} \right\}_{f,i,p \in Q_j}$
Glints	$(\sigma^G)^2$	$a_\sigma + \frac{M}{2}$	$\frac{\text{trace}(\mathbf{\Gamma}_\rho^{-1} \mathbf{W} \mathbf{W}^H)}{2}$	$M = \sum_{f,p} \delta_f^{G,(p)}, \mathbf{W} = \text{vec} \left\{ \underline{g}_{f,i}^{(p)} \right\}_{\{p,f,i: \delta_f^{G,(p)}=1\}}$
Target	$(\sigma^M)^2$	$a_\sigma + \frac{M}{2}$	$\frac{\text{trace}(\mathbf{W} \mathbf{W}^H)}{2}$	$M = \sum_{f,i,p} \delta_{f,i}^{M,(p)}, \mathbf{W} = \text{vec} \left\{ \underline{m}_{f,i}^{(p)} \right\}_{\{p,f,i: \delta_{f,i}^{M,(p)}=1\}}$
Calibration	$(\sigma^H)^2$	$a_\sigma + \frac{M}{2}$	$\frac{\text{trace}(\mathbf{W} \mathbf{W}^H)}{2}$	$M = KFNP/ Z_g , \mathbf{W} = \text{vec} \left\{ z_{k,f,i}(g) \right\}_{k,f,i,g}$
Additive noise	$(\sigma^V)^2$	$a_\sigma + \frac{M}{2}$	$\frac{\text{trace}(\mathbf{W} \mathbf{W}^H)}{2}$	$M = KFNP, \mathbf{W} = \text{vec} \left\{ \frac{i_{k,f,i}^{(p)}}{h_{k,f,i}^{(p)}} - l_{k,f,i}^{(p)} - s_{k,f,i}^{(p)} \right\}_{k,f,i,p}$

Note: The set  $Q_j$  is defined as  $Q_j \triangleq \{p : c^{(p)} = j\}$



Note that this is in the form of the Multivariate-Normal-Inverse-Wishart conjugate distribution given  $\rho$ . This leads to the well known posterior parameters:

$$\Gamma_\rho \left( \tilde{\mathbf{W}}, \sigma^2, \rho \right) \sim \text{InvWishart}(\Lambda_\rho a_\Gamma + \sum_{m=1}^n \left( \underline{y}_m - \tilde{\boldsymbol{\mu}} \right) \left( \underline{y}_m - \tilde{\boldsymbol{\mu}} \right)^H, \nu_\Gamma + n) \quad (101)$$

where  $\Lambda_\rho = \rho \mathbf{1}_K \mathbf{1}_K^T + (1 - \rho) \mathbf{I}_{K \times K}$ . Ideally, we would like to sample both  $\Gamma_\rho$  and  $\rho$  jointly. Even though we can simply sample from equation (101), the same is not true for the density

$$\rho \left( \tilde{\mathbf{W}}, \sigma^2, \rho \right) \quad (102)$$

which is required in order to jointly sample these parameters. Fortunately, we know that

$$f(\Gamma_\rho, \rho | \tilde{\mathbf{W}}, \sigma^2) \propto f(\Gamma_\rho | \tilde{\mathbf{W}}, \sigma^2, \rho) f(\rho) \quad (103)$$

which is easily evaluated since we have closed form functions for both of these densities. Thus, we can use Metropolis-Hastings to sample  $\rho$  and  $\Gamma_\rho$ .

2) *Indicator probabilities:* In the basic model where the indicator Beta distribution parameters do not depend spatially or temporally, then the posterior indicator probabilities for

$$\delta \sim \text{Bernoulli}(\pi) \quad (104)$$

$$\pi \sim \text{Beta}(a_\pi, b_\pi) \quad (105)$$

are given by

$$\pi | \delta \sim \text{Beta}(a_\pi + \delta, b_\pi + (1 - \delta)) \quad (106)$$

Note that we can modify  $a_\pi$  and  $b_\pi$  as in Section IV-A. However, the posterior inference for the probabilities is identical by replacing  $a_\pi$  and  $b_\pi$  by their spatiotemporally varying version.

3) *Object class probabilities:* We use a Multinomial-Dirichlet conjugate pair to determine object class assignments, where the class probabilities  $\underline{q}$  have a prior Dirichlet distribution with  $c_j = 1/J$  for  $j = 1, 2, \dots, J$ . Then, after observing the class assignments, we can calculate the number of pixels in any class

$$N_j = |Q_j| = |\{p : c^{(p)} = j\}| \quad (107)$$

Then the posterior distribution for the class probabilities is given by

$$\underline{q} | \{N_j\}_{j=1}^J \sim \text{Gamma}(\underline{q} + \underline{N}, 1) / J \quad (108)$$

where  $[\underline{N}]_j = N_j$ .

## REFERENCES

- [1] J. Fienup, "Detecting moving targets in SAR imagery by focusing," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 37, no. 3, pp. 794–809, 2001.
- [2] R. Deming, "Along-track interferometry for simultaneous sar and gmti: application to gotcha challenge data," in *Proceedings of SPIE*, vol. 8051, 2011, p. 80510P.
- [3] R. Deming, S. MacIntosh, and M. Best, "Three-channel processing for improved geo-location performance in sar-based gmti interferometry," in *Proceedings of SPIE*, vol. 8394, 2012, p. 83940F.
- [4] K. Ranney and M. Soumekh, "Signal subspace change detection in averaged multilook sar imagery," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 44, no. 1, pp. 201–213, 2006.
- [5] M. Soumekh, "Moving target detection in foliage using along track monopulse synthetic aperture radar imaging," *Image Processing, IEEE Transactions on*, vol. 6, no. 8, pp. 1148–1163, Aug. 1997.
- [6] J. Ender, "Space-time processing for multichannel synthetic aperture radar," *Electronics & Communication Engineering Journal*, vol. 11, no. 1, pp. 29–38, 1999.
- [7] J. Jao, "Theory of synthetic aperture radar imaging of a moving target," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 9, pp. 1984–1992, 2001.
- [8] B. Borden and M. Mumford, "A statistical glint/radar cross section target model," *Aerospace and Electronic Systems, IEEE Transactions on*, no. 5, pp. 781–785, 1983.

- [9] F. Posner, "Texture and speckle in high resolution synthetic aperture radar clutter," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 31, no. 1, pp. 192–203, 1993.
- [10] R. Raney and G. Wessels, "Spatial considerations in sar speckle consideration," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 26, no. 5, pp. 666–672, 1988.
- [11] L. Borcea, T. Callaghan, and G. Papanicolaou, "Synthetic aperture radar imaging and motion estimation via robust principle component analysis," *arXiv preprint arXiv:1208.3700*, 2012.
- [12] A. Khwaja and J. Ma, "Applications of compressed sensing for sar moving-target velocity estimation and image compression," *Instrumentation and Measurement, IEEE Transactions on*, vol. 60, no. 8, pp. 2848–2860, 2011.
- [13] C. Gierull, "Statistical analysis of multilook sar interferograms for cfar detection of ground moving targets," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 42, no. 4, pp. 691–701, 2004.
- [14] Y. Bazi, L. Bruzzone, and F. Melgani, "An unsupervised approach based on the generalized gaussian model to automatic change detection in multitemporal sar images," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 43, no. 4, pp. 874–887, 2005.
- [15] F. Bovolo and L. Bruzzone, "A detail-preserving scale-driven approach to change detection in multitemporal sar images," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 43, no. 12, pp. 2963–2972, 2005.
- [16] R. Perry, R. Dipietro, and R. Fante, "Sar imaging of moving targets," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 35, no. 1, pp. 188–200, 1999.
- [17] S. Zhu, G. Liao, Y. Qu, Z. Zhou, and X. Liu, "Ground moving targets imaging algorithm for synthetic aperture radar," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 49, no. 1, pp. 462–477, 2011.
- [18] B. Guo, D. Vu, L. Xu, M. Xue, and J. Li, "Ground moving target indication via multichannel airborne sar," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 49, no. 10, pp. 3753–3764, oct. 2011.
- [19] J. Guo, Z. Li, and Z. Bao, "Adaptive clutter suppression and resolving of velocity ambiguities for an experimental three-channel airborne synthetic aperture radar-ground moving target indication system," *Radar, Sonar & Navigation, IET*, vol. 5, no. 4, pp. 426–435, 2011.
- [20] J. Wright, A. Ganesh, S. Rao, Y. Peng, and Y. Ma, "Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization," *submitted to Journal of the ACM*, 2009.
- [21] Z. Lin, M. Chen, and Y. Ma, "The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices," *arXiv preprint arXiv:1009.5055*, 2010.
- [22] E. Candes, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *Journal of the ACM*, vol. 58, no. 3, 2011.
- [23] X. Ding, L. He, and L. Carin, "Bayesian robust principal component analysis," *Image Processing, IEEE Transactions on*, vol. 20, no. 12, pp. 3419–3430, 2011.
- [24] M. Tipping, "Sparse Bayesian learning and the relevance vector machine," *The Journal of Machine Learning Research*, vol. 1, pp. 211–244, 2001.
- [25] D. P. Wipf and B. D. Rao, "Sparse Bayesian learning for basis selection," *IEEE Transactions on Signal Processing*, vol. 52, no. 8, pp. 2153–2164, August 2004.
- [26] G. E. Newstadt, E. Zelnio, L. Gorham, and A. O. H. III, "Detection/tracking of moving targets with synthetic aperture radars," E. G. Zelnio and F. D. Garber, Eds., vol. 7699, no. 1. SPIE, 2010, p. 76990I. [Online]. Available: <http://link.aip.org/link/?PSI/7699/76990I/1>



**Gregory Newstadt** received the B.S. degrees (summa cum laude) from Miami University, Oxford, OH, in 2007 in Electrical Engineering and in Engineering Physics. He also received the M.S.E in Electrical Engineering: Systems (2009), M.A. in Statistics (2012) and Ph.D. in Electrical Engineering: Systems (2013) degrees from the University of Michigan, Ann Arbor, MI.

He is currently a postdoctoral researcher and lecturer at the University of Michigan, Ann Arbor, MI, in Electrical Engineering (Systems). His research interests include detection, estimation theory, target tracking, sensor fusion, and statistical signal processing.



**Edmund Zelnio** graduated from Bradley University, Peoria, Illinois, in 1975 and has pursued doctoral studies at The Ohio State University in electromagnetics and at Wright State University in signal processing. He has had a 37 year career with the Air Force Research Laboratory (AFRL), Wright Patterson AFB, Ohio where he has spent 35 years working in the area of automated exploitation of imaging sensors primarily addressing synthetic aperture radar. He is a former division chief and technical advisor of the Automatic Target Recognition Division of the Sensors Directorate in AFRL and serves in an advisory capacity to the Department of Defense and the intelligence community. He is currently the director of the Automatic Target Recognition Center in AFRL. He is the recipient of the 53rd DoD Distinguished Civilian Service Award and is a fellow of the Air Force Research Laboratory.



**Alfred O. Hero, III** received the B.S. (summa cum laude) from Boston University (1980) and the Ph.D from Princeton University (1984), both in Electrical Engineering. Since 1984 he has been with the University of Michigan, Ann Arbor, where he is the R. Jamison and Betty Professor of Engineering. His primary appointment is in the Department of Electrical Engineering and Computer Science and he also has appointments, by courtesy, in the Department of Biomedical Engineering and the Department of Statistics. In 2008 he was awarded the the Digiteo Chaire d'Excellence, sponsored by Digiteo Research Park in Paris, located at the Ecole Supérieure d'Electricité, Gif-sur-Yvette, France. He has held other visiting positions at LIDS Massachusetts Institute of Technology (2006), Boston University (2006), I3S University of Nice, Sophia-Antipolis, France (2001), Ecole Normale Supérieure de Lyon (1999), Ecole Nationale Supérieure des Télécommunications, Paris (1999), Lucent Bell Laboratories (1999), Scientific Research Labs of the Ford

Motor Company, Dearborn, Michigan (1993), Ecole Nationale Supérieure des Techniques Avancées (ENSTA), Ecole Supérieure d'Electricité, Paris (1990), and M.I.T. Lincoln Laboratory (1987 - 1989).

Alfred Hero is a Fellow of the Institute of Electrical and Electronics Engineers (IEEE). He has been plenary and keynote speaker at major workshops and conferences. He has received several best paper awards including: a IEEE Signal Processing Society Best Paper Award (1998), the Best Original Paper Award from the Journal of Flow Cytometry (2008), and the Best Magazine Paper Award from the IEEE Signal Processing Society (2010). He received a IEEE Signal Processing Society Meritorious Service Award (1998), a IEEE Third Millennium Medal (2000) and a IEEE Signal Processing Society Distinguished Lecturership (2002). He was President of the IEEE Signal Processing Society (2006-2007). He sits on the Board of Directors of IEEE (2009-2011) where he is Director Division IX (Signals and Applications).

Alfred Hero's recent research interests have been in detection, classification, pattern analysis, and adaptive sampling for spatio-temporal data. Of particular interest are applications to network security, multi-modal sensing and tracking, biomedical imaging, and genomic signal processing.